Study Protocol

P4-C1-022

P4-C2-015

P4-C2-016

DARWIN EU® - Characterisation of data capture related to suicidality and depression across the DARWIN EU® network

05/02/2025

Version 4.0

Authors: Nicholas Hunt, Guido van Leeuwen, Katia Verhamme

Public

CONTENTS

Study title[1]	DARWIN EU® - Characterisation of data capture related to suicidality and depression across the DARWIN EU® network
Protocol version	V4.0
Date	05/02/2025
EUPAS number	EUPAS1000000825
Active Substance	None
Medicinal Product	None
Research question and objectives	Research questions 1. What proportion of individuals within the DARWIN EU® network have a record of suicide, suicide attempt, suicidal ideation, self-harm, depression, symptom measurement data, procedures, and depression resolution? 2. What is the median number, rate, and changes in rate over time of suicide, suicide-related events, and depression per individual within the DARWIN EU® network? 3. What are the characteristics of individuals with suicide, suicide-related events and depression? Objectives 1. To estimate the incidence and prevalence of individuals with a record of suicide, suicide-related events, depression, healthcare utilisation, symptom measurement data, procedures, and depression remission within the DARWIN EU® network. 2. To characterise reporting of suicide, suicide-related events, depression, symptom measurement, healthcare utilisation, and resolution reporting within the DARWIN EU® network in terms of: a) Mean and median of number of event reporting per individual within the study period. b) The mean cumulative function of event recording c) Describe the values of depression symptom measurement data. 3. To describe the characteristics of individuals with records of suicide, suicide-related events, and depression in terms of demographic characteristics, concomitant medications, comorbidities, procedures, and lifestyle factors.
Countries of study	Croatia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Norway, Portugal, Spain, Sweden, The Netherlands, and The United Kingdom
Authors	Nicholas Hunt: n.hunt@darwin-eu.org Guido van Leeuwen: g.vanleeuwen@darwin-eu.org Katia Verhamme: k.verhamme@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/terms	Description
ATC	Anatomical Therapeutic Chemical
CBT	Cognitive behavioural therapy
CDM	Common Data Model
CC	Coordination centre
CPRD GOLD	CPRD GOLD
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
DQ	Data quality
DQD	Data Quality Dashboard
EBB	Estonian Biobank
ECT	Electro compulsive therapy
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
GDPR	General Data Protection Regulation
GP	General Practitioner
ICD	International Classification of Diseases
InGef RDB	InGef - Institute for Applied Health Research Berlin GmbH
IP	Inpatient
IPT	Interpersonal therapy
IPCI	Integrated Primary Care Information
IQR	Interquartile range
MCF	Mean cumulative function
NAJS	Croatian National Public Health Information System
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
PHQ-9	Patient Healthcare Questionnaire-9
rTMS	Repetitive transcranial magnetic stimulation
RxNorm	Medical prescription normalized
RWD	Real World Data
SNOMED	Systematized Nomenclature of Medicine
tDCS	Transcranial direct current stimulation
VNS	Vagus nerve stimulation
WHO	World Health Organisation

1. TITLE

DARWIN EU® - Characterisation of data capture related to suicidality and depression across the DARWIN EU® network

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Nicholas Hunt Guido van Leeuwen Katia Verhamme	Erasmus MC Erasmus MC Erasmus MC
Data Scientist	Ger Inberg Maarten van Kessel	Erasmus MC Erasmus MC
Study Manager	Natasha Yefimenko Nosova	Erasmus MC
Data source	**Names**	**Data Partner Organisation****
P4-C1-022*		
Croatia: Croatian National Public Health Information System (NAJS)	Anamaria Jurčević Antea Jezidžić Jakov Vuković Ivan Pristaš	Croatian Institute of Public Health
Denmark: Danish Data Health Registries (DK-DHR)	Elvira Bräuner Susanne Bruun	Danish Medicines Agency (DKMA)
Estonia: Estonian Biobank (EBB)	Marek Oja Raivo Kolde	University of Tartu
Germany: InGef Research Database (InGef RDB)	Raeleesha Norris Josephine Jacob Annika Vivirito Alexander Harms	InGef - Institute for Applied Health Research Berlin GmbH
The Netherlands: Integrated Primary Care Information (IPCI)	Katia Verhamme Mees Mosseveld Guido van Leeuwen	Erasmus MC
The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)	Antonella Delmestri Marta Pineda Moncusí	University of Oxford
P4-C2-015*		
Greece: Papageorgiou General Hospital (PGH)	Alexandros Rekkas Achilleas Chytas Anastasia Farmaki Pantelis Natsiavas Loukas Athanasiadis Eleni Parlapani Ioannis Gliatas Patroklos Theocharis	Papageorgiou General Hospital

	Anastasia Lavrentiadou Panagiota Karamouzi Vasiliki Holeva	
Portugal: Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)	Joao Firmino Machado Ana Pinto Tiago Taveira Mesquita Bastos Firmino Machado Teresa Monjardino	Clinical Academic Center Egas Moniz Health Alliance
Portugal: Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA)	Joao Firmino Machado Ana Pinto Tiago Taveira	Clinical Academic Center Egas Moniz Health Alliance
Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)	Miguel Jesus Gil Garcia María del Mar Martín Pérez Virginia Arroyo Nebreda Hermenegildo Carlos Martínez-Alcalá García Ana Llorente Garcia Miguel Angel Macia Martinez	La Agencia Española de Medicamentos y Productos Sanitarios
Spain: Hospital Universitario 12 de Octubre (H12O)	Juan Luis Cruz Bermudez Noelia Garcia Barrio Paula Rubio Mayo	Hospital Universitario 12 de Octubre
Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)	Juan Manuel Ramírez Anguita Miguel Angel Mayer Pujadas Angela Leis	Parc de Salut Mar Barcelona
Spain: The Information System for Research on Primary Care (SIDIAP)	Agustina Giuliodori Picco Elena Roel Herranz Irene López Sánchez Laura Granés González Talita Duarte Salles	L'Institut d'Investigació en Atenció Primària de Salut Jordi Gol
Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)	Huiqi Li Fredrik Nyberg	Swedish Medical Products Agency - Gothenburg University
P4-C2-016*		
Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)	Eric Fey Tiina Wahlfors Kimmo Porkka Marianna Niemi	FinOMOP
Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)	Tiina Wahlfors Hakkarainen Leena Kati Kristiansson	FinOMOP

	Sampo Kukkurainen	
Finland: Finnish Care Register for Health Care (FinOMOP-THL)	Tiina Wahlfors Toni Lehtonen Gustav Klingstedt	FinOMOP
France: Assistance publique Hôpitaux de Marseille (APHM)	Laurent Boyer Dorian Grousset Vanessa Pauly	Assistance publique Hôpitaux de Marseille
Hungary: Semmelweis University Clinical Data (SUCD)	Ágota Mészáros Bagyura Zsolt István Kiss Loretta Zsuzsa Héja Tibor	Semmelweis University
Norway: Norwegian Linked Health Registry data (NLHR)	Hedvig Nordeng Saeed Hayati	University of Oslo
Portugal: Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT)	Fernando Montenegro Sá Nuno Silva	Unidade Local de Saúde de Matosinhos
Spain: Platafoma de Recerca en Informació Sanitària de les Illes Balears (PRISIB)	Pau Pericas Pulido Joan Vicenç Cladera Salva	Health Data Research Platform of the Balearic Islands
Spain: Valencia Health System Integrated Dataset (VID)	Gabriel Sanfélix Gimeno Celia Robles Cabaniñas Fran Llopis Cardona	Fundació per al Foment de la Investigació Sanitària i Biomèdica de la Comunitat Valenciana

*Three DARWIN EU® studies are to be performed (P4-C1-022, P4-C2-015, P4-C2-016) to account for a total of 23 data sources.
*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® - Characterisation of data capture related to suicidality and depression across the DARWIN EU® network.

Rationale and background

Depression is a prevalent health condition causing significant healthcare utilisation, morbidity, and mortality. Self-harm, suicidal ideation, attempt, and suicide also comprise a complex and significant public health issue. Safety concerns on the use of medicines in these populations or medicines as a cause of these conditions are increasingly common. A better understanding on how well data on these areas are captured will be important to inform the feasibility of conducting Real world data (RWD) studies on these populations or their use as outcomes in DARWIN EU® studies.

Objectives

1. To estimate the incidence and prevalence of individuals with a record of suicide, suicide-related events, depression, healthcare utilisation, symptom measurement data, procedures, and depression remission within the DARWIN EU® network.
2. To characterise reporting of suicide, suicide-related events, depression, symptom measurement, healthcare utilisation, and resolution reporting within the DARWIN EU® network in terms of:
   a) Mean and median of number of event reporting per individual within the study period.
   b) The mean cumulative function of event recording
   c) Describe the values of depression symptom measurement data.
3. To describe the characteristics of individuals with records of suicide, suicide-related events, and depression in terms of demographic characteristics, concomitant medications, comorbidities, procedures, and lifestyle factors.

Methods

Study design

- A descriptive study including the general population will be conducted to address objectives 1 and 2c.
- A descriptive study including individuals with records of the outcomes of interest will be conducted to address objectives 2a, 2b, and 3.

Population

For objective 1 and 2c, the study population will include all individuals present in the data source during the study period 01/01/2015 to 31/12/2024 (or to the end of available data) and with at least 365 days of data source history prior to index date. For objective 2a, 2b, and 3, the study population will include individuals with a first occurrence of suicide (completed), suicide attempt, suicide ideation, self-harm, or depression in the study period with at least 365 days of prior observation.

Variables

Outcome:

Suicide (completed), composite suicide-related events (suicide, suicide attempt, suicide ideation, self-harm), composite fatal suicide-related events, composite non-fatal suicide-related events, suicide attempt, suicidal ideation, self-harm, depression, symptom measurement data (Patient Health Questionnaire-9 (PHQ9) depression scale), healthcare utilisation, electrotherapy, and psychotherapy.

Relevant covariates:

Demographic characteristics, drug prescriptions (antidepressants, antipsychotics, benzodiazepines, stimulants, hypnotics), conditions (schizophrenia, anxiety disorder, substance use disorder, personality disorders), procedures (psychotherapy, electrotherapy), and observation occurrences (smoking, obesity).

Data sources*

P4-C1-022:

1. Croatia: Croatian National Public Health Information System (NAJS)
2. Denmark: Danish Data Health Registries (DK-DHR)
3. Estonia: Estonian Biobank (EBB)
4. Germany: InGef Research Database (InGef RDB)
5. The Netherlands: Integrated Primary Care Information (IPCI)
6. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

P4-C2-015:

1. Greece: Papageorgiou General Hospital (PGH)
2. Portugal: Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)
3. Portugal: Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA)
4. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
5. Spain: Hospital Universitario 12 de Octubre (H12O)
6. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
7. Spain: The Information System for Research on Primary Care (SIDIAP)
8. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

P4-C2-016:

1. Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)
2. Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)
3. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
4. France: Assistance publique Hôpitaux de Marseille (APHM)
5. Hungary: Semmelweis University Clinical Data (SUCD)
6. Norway: Norwegian Linked Health Registry data (NLHR)
7. Portugal: Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT)
8. Spain: Platafoma de Recerca en Informació Sanitària de les Illes Balears (PRISIB)
9. Spain: Valencia Health System Integrated Dataset (VID)

*Three DARWIN EU® studies are to be performed (P4-C1-022, P4-C2-015, P4-C2-016) to account for a total of 23 data sources.

Statistical analysis

Characteristics will be described by means of pre-specified characterisation. Covariates of interest will be reported as counts and proportions. Yearly incidence rates per 100,000 person-years and period prevalence (proportion) of the outcomes will be estimated in the general population, overall and stratified by age categories and sex. Incidence rates will be given together with 95% Poisson confidence intervals. To assess the sequence of events, the mean cumulative function at time t will report the cumulative number of events per person on average up to time t, and we will estimate the mean and median number of events per individual. The statistical analyses will be performed based on Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)-mapped data using *IncidencePrevalence* and *CohortCharacteristics* R packages.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	January 2026
Creation of Analytical code	January 2026
Execution of Analytical Code on the data	February 2026
Draft Study Report	March 2026
Final Study Report	April 2026

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

6. RATIONALE AND BACKGROUND

Depression is a prevalent health condition causing significant healthcare utilisation, morbidity, and mortality. Self-harm, suicidal ideation, attempt, and suicide also comprise a complex and significant public health issue. Safety concerns on the use of medicines in these populations or medicines as a cause of these conditions are increasingly common. Real world data offer the potential to provide important safety and contextual insights in this area, but its utility depends on the quality of related data, including an understanding of: extensiveness (is there sufficient data on suicide and suicide-related events in the DARWIN EU® network), coherence (can the data be used for analyses on research questions relating to suicide and suicide-related events), reliability, relevance, and timeliness. Several DARWIN EU® studies have been executed investigating suicide or suicide-related events as the outcome, however the extent of underreporting potential misclassification of suicide remains uncertain.[1, 2]

The aim of this project is to characterise individuals with a record of suicide, suicidal attempt and ideation, self-harm, and depression related data mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) in data sources included in the DARWIN EU® network. The characterisation in this study includes outcomes and covariates such as comedications, comorbidities, symptom measurements, procedures and how longitudinal changes and possible repeating events (e.g., ideation, attempt) may be captured. A better understanding on how well data on these areas are captured will be important to inform the feasibility of conducting RWD studies on these populations or their use as outcomes in DARWIN EU® studies. This study is part of three studies, which include one off-the-shelf (P4-C1-022) and two routine repeated studies (P4-C2-015 and P4-C2-16), that will account for 23 DARWIN EU® data sources.

7.

8. RESEARCH QUESTION AND OBJECTIVES

Research questions

1. What proportion of individuals within the DARWIN EU® network have a record of suicide, suicide attempt, suicidal ideation, self-harm, depression, symptom measurement data, procedures, and suicide and depression resolution?
2. What is the median number, rate, and changes in rate over time of suicide (related) events and depression per individual within the DARWIN EU® network?
3. What are the characteristics of individuals with suicide, suicide-related events and depression?

Research objectives

1. To estimate the incidence and prevalence of individuals with a record of suicide, suicide-related events, depression, healthcare utilisation, symptom measurement data, procedures, and depression remission within the DARWIN EU® network.
2. To characterise reporting of suicide, suicide-related events, depression，symptom measurement, healthcare utilisation, and resolution reporting within the DARWIN EU® network in terms of:
    a. Mean and median of number of event reporting per individual within the study period.
    b. The mean cumulative function of event recording
    c. Describe the values of depression symptom measurement data.
3. To describe the characteristics of individuals with records of suicide, suicide-related events, and depression in terms of demographic characteristics, concomitant medications, comorbidities, procedures, and lifestyle factors.

9. RESEARCH METHODS

9.1. Study design

The retrospective cohort study will comprise of:

- A descriptive epidemiology study including the general population will be conducted to address objective 1 and 2c.
- A descriptive study including individuals with the events of interest (suicide, suicide attempt, suicide ideation, self-harm, or depression) will be conducted to address objective 2a, 2b, and 3.

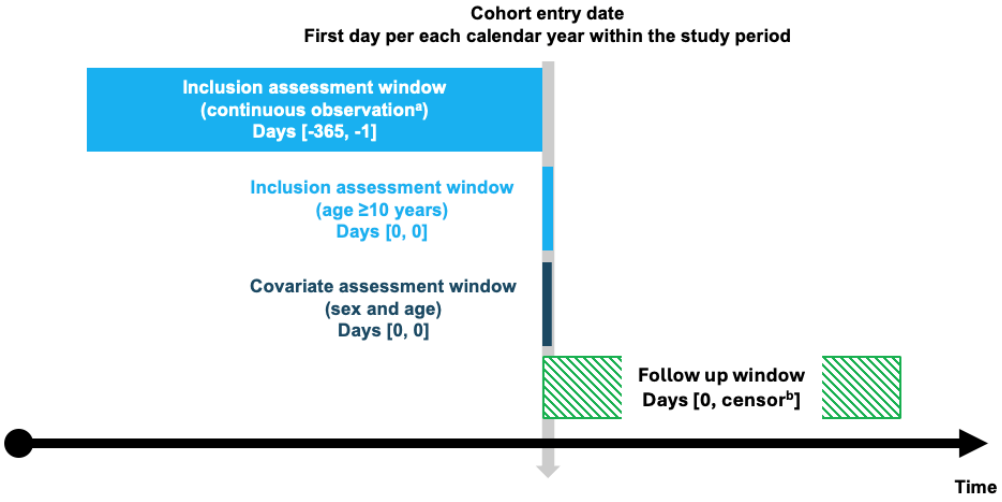

Figure 1. Graphical depiction of the study design of objective 1 and 2c.

a. Does not apply to individuals in hospital data sources
b. Death, disenrollment, end of data source availability, end of each calendar year (i.e., 31st December), or end of the study period (31/12/2024)

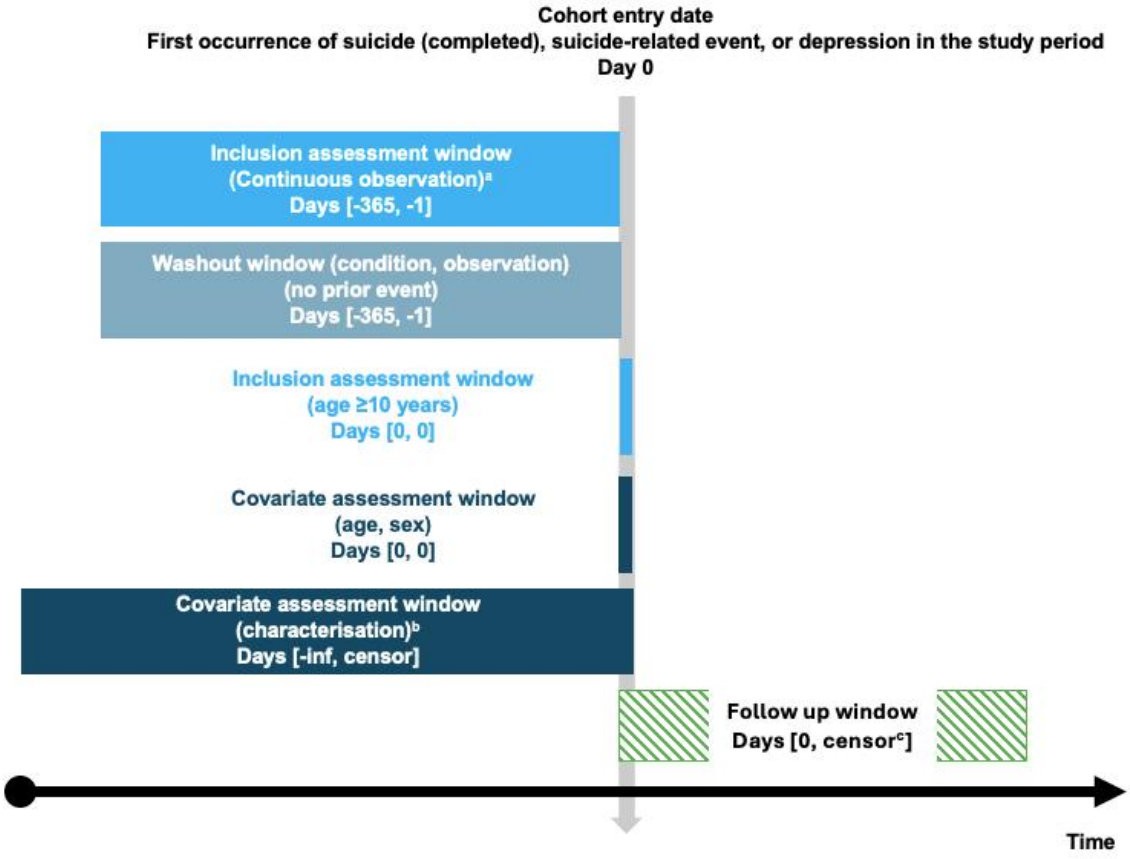

Figure 2. Graphical depiction of the study design of objective 2a, 2b, and 3.

a. Does not apply to individuals in hospital data sources
b. Conditions, drug prescriptions, observations, and procedure occurrences (defined in **Section 8.6.3**)
c. Death, disenrollment, end of data source availability, five years, or end of the study period (31/12/2024)

9.2. Follow-up

For objective 1, follow-up will start on the respective date of the latest of the following: i) study start date 01/01/2015, ii) first date in each calendar year, iii) date at which individuals have 365 days of prior history, or iv) the date of 10th birthday. End of follow-up will be defined as the earliest of loss to follow-up, death, or end of observation period (the latest available data), end of each calendar year, whichever occurs first.

For objective 1, prevalence requires an appropriate denominator population and contributes observation time to first be identified. Study participants in the denominator population will begin contributing time to the when eligible for inclusion (**Figure 1**) and will be followed until end of observation, as seen in **Figure 3**.

In this example, person ID 1 already has sufficient prior history before the study start date, and the observation period ends after the study end date, so this individual will contribute during the complete study period. Person IDs 2 and 4 enter the study only when they have sufficient prior history. Person ID 3 leaves when exiting the data source (the end of the observation period). Lastly, person ID 5 has two observation periods in the data source. The first period contributes time from the study start until the end of the observation period. The second starts contributing time again once sufficient prior history is reached and exits at the study end date.

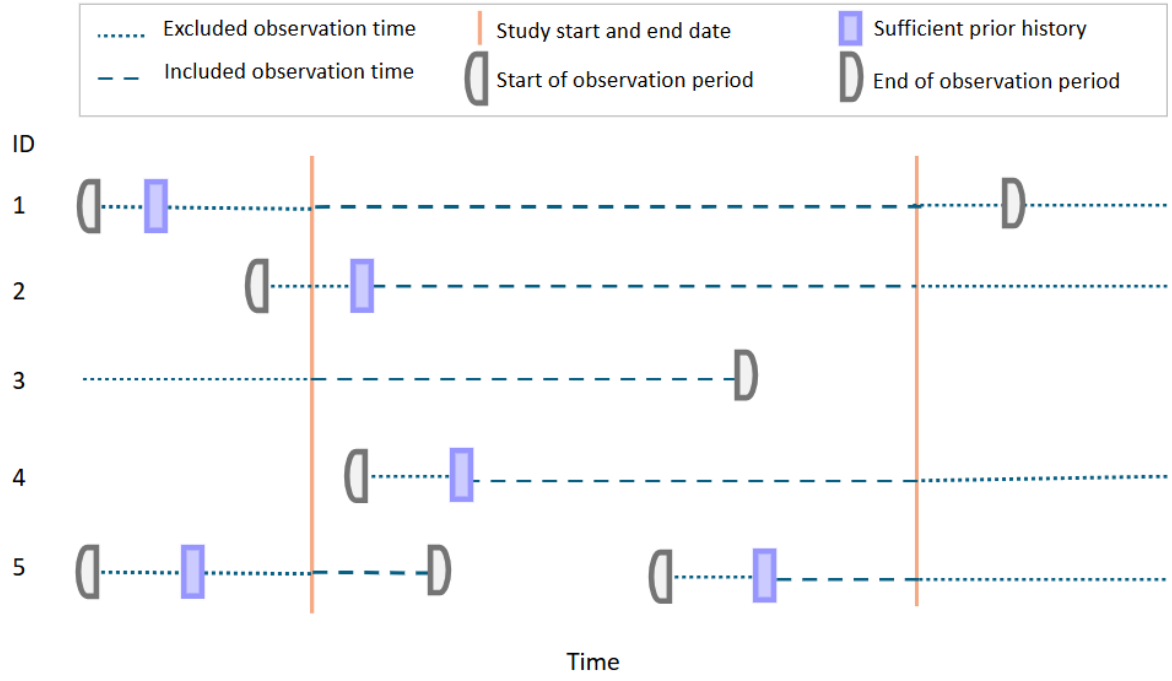

Figure 3. Included observation time for the denominator population.

9.3. Study population with inclusion and exclusion criteria

For **objective 1 and 2c**, the general population will constitute the study population:

Inclusion criteria

- Minimum 365 days of available history before index date (except for hospital data sources)
- Aged ≥10 years of age. This age cutoff is based on the World Health Organisation (WHO) definition for adolescence (age 10 to 19), since we do not expect relevant results in lower age groups.[3]

For **objective 2a, 2b. and 3**:

Inclusion criteria

- Minimum 365 days of available history before index date (except for hospital data sources)
- Aged ≥10 years of age
- Condition or observation of completed suicide, suicide-related events (suicide attempt, suicide ideation, or self-harm), or depression

9.4. Study setting and data sources

This study will be conducted using routinely collected data from 23 data sources in the DARWIN EU® network of data partners from 14 European countries, of which 12 EU member states. All data were *a priori* mapped to the OMOP CDM.

Table 1. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to
P4-C1-022*						
Croatia	Croatian National Public Health Information System (NAJS)	Primary and secondary care	Registry	4,295,300	01/01/2018–31/12/2024	Objective 1, 2, and 3
Denmark	Danish Data Health Registries (DK-DHR)	Secondary care	Registry	5,984,000	01/01/2015–31/11/2024	Objective 1, 2, and 3
Estonia	Estonian Biobank (EBB)	Primary and secondary care	Biobank	212,300	01/01/2015–31/12/2023	Objective 1, 2, and 3
Germany	InGef Research Database (InGef RDB)	Primary and secondary care	Claims	7,665,200	01/01/2016–31/12/2024	Objective 1, 2, and 3
The Netherlands	Integrated Primary Care Information (IPCI)	Primary care	EHR	1,332,500	01/01/2015–31/12/2024	Objective 1, 2, and 3
The United Kingdom	Clinical Practice Research Datalink GOLD (CPRD GOLD)	Primary care	EHR	2,632,700	01/01/2015–31/12/2024	Objective 1, 2, and 3
P4-C2-015*						
Greece	PGH	Secondary care	EHR	55.7k	01/01/2015–31/11/2024	Objective 1, 2 (except 2b), and 3

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to
Portugal	EMDB-ULSGE	Primary and secondary care	EHR	130k	01/01/2015–31/11/2024	Objective 1, 2, and 3
Portugal	EMDB-ULSRA	Primary and secondary care	EHR	89.5k	01/01/2015–31/11/2024	Objective 1, 2, and 3
Spain	BIFAP	Primary and secondary care (inpatient)	EHR	23.2m	01/01/2015–31/11/2024	Objective 1, 2, and 3
Spain	H12O	Secondary care	EHR	295k	01/01/2015–31/11/2024	Objective 1, 2 (except 2b), and 3
Spain	IMASIS	Secondary care	EHR	103k	01/01/2015–31/11/2024	Objective 1, 2 (except 2b), and 3
Spain	SIDIAP	Primary and secondary care	EHR	5.95m	01/01/2015–31/11/2024	Objective 1, 2, and 3
Sweden	HI-SPEED	Primary and secondary care	Registry	10.6m	01/01/2016–31/11/2024	Objective 1, 2, and 3
P4-C2-016*						
Finland	FinOMOP-HUS	Secondary care	EHR	616k	01/01/2015–31/11/2024	Objective 1, 2 (except 2b), and 3
Finland	FinOMOP-TaUH Pirha	Secondary care	EHR	263k	01/01/2015–31/11/2024	Objective 1, 2 (except 2b), and 3
Finland	FinOMOP-THL	Primary and secondary care	EHR, registry	5.7m	01/01/2015–31/11/2024	Objective 1, 2, and 3
France	APHM	Secondary care	EHR, Claims, registry	250k	01/01/2015–31/11/2024	Objective 1, 2 (except 2b), and 3
Hungary	SUCD	Secondary care	EHR, Claims, registry	227k	01/01/2015–31/11/2024	Objective 1, 2 (except 2b), and 3

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to
Norway	NLHR	Primary and secondary care	Registry	5.55m	01/01/2015–31/11/2024	Objective 1, 2, and 3
Portugal	ULSM-RT	Primary and secondary care	EHR	76k	01/01/2015–31/11/2024	Objective 1, 2, and 3
Spain	PRISIB	Primary and secondary care	EHR, registry	1.79m	01/01/2015–31/11/2024	Objective 1, 2, and 3
Spain	VID	Primary and secondary care	EHR, registry, other	5.58m	01/01/2015–31/11/2024	Objective 1, 2, and 3

*Three DARWIN EU® studies are to be performed (P4-C1-022, P4-C2-015, P4-C2-016) to account for a total of 23 data sources. Acronyms: NAJS = Croatian National Public Health Information System, DK-DHR = Danish Data Health Registries, EBB = Estonian Biobank, FinOMOP THL = Finnish Care Register for Health Care, FinOMOP HUS = Hospital District of Helsinki and Uusimaa, FinOMOP-TaUH Pirha = Tampere University Hospital patient cohort, APHM = Assistance Publique Hôpitaux de Marseille, InGef = Institute for Applied Health Research Berlin GmbH Research Database, PGH = Papageorgiou General Hospital, SUCD = Semmelweis University Clinical Data, IPCI = Integrated Primary Care Information, EHR = Electronic Health Records, NLHR = Norwegian Linked Health Registry data, EMDB-ULSGE = Egas Moniz Health Alliance database - Gaia E Espinho, EMDB-ULSRA = Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro), ULSM-RT = Unidade Local de Saúde de Matosinhos Realtime Database, BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, H12O = Hospital Universitario 12 de Octubre, IMASIS = Institut Municipal Assistència Sanitària Information System, PRISB = Platafoma de Recerca en Informació Sanitària de les Illes Balears, SIDIAP = The Information System for the Development of Research in Primary Care, VID = Valencia Health System Integrated Dataset, HI-SIPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD = Clinical Practice Research Datalink GOLD

Data sources selection

The purpose of this study is data characterisation, so rather than a selection of data sources for a specific study based on known elements (i.e., more complete data), we will include a broader range of data sources. These data sources fulfil the criteria required in terms of data quality, completeness, timeliness, and representativeness for disease epidemiology and characterisation studies in general, while covering different regions of Europe (**Annex II**).

9.5. Study period

The study period is from 01/01/2015 to 31/12/2024 or the most recent data available for each contributing data source. The start of observation in InGef RDB and HI-SPEED is within the study period, therefore the availability of data with 365 days prior history commences from 01/01/2016. In NAJS, the availability of data begins on 01/01/2018. Therefore, due to data availability these data sources will have a later study period start date than 01/01/2015.

9.6. Variables

9.6.1. Exposure

There are no medications specifically assessed under this study. Characterisation by concomitant medications is described in **Section 8.6.4.**

9.6.2. Outcome

For Objective 1

Prevalence of the following variables will be estimated using standard concept IDs applicable to each domain of variable. The outcomes for this objective are as follows:

- Observation occurrence of suicide (completed)
- Composite suicide-related events (suicide, suicide attempt, suicide ideation, self-harm)
- Composite fatal suicide-related events (suicide, suicide attempt, suicide ideation, self-harm) with a death record within ±30 days
- Composite non-fatal suicide-related events (suicide, suicide attempt, suicide ideation, self-harm) without a death record within ±30 days
- Observation occurrence of suicide attempt
- Observation occurrence of suicidal ideation
- Observation occurrence of intentional self-harm
- Condition occurrence of depression
- Measurement occurrence of Patient Health Questionnaire-9 (PHQ9) depression scale, Beck Depression Inventory (I and II), Hamilton Depression Scale, Columbia Suicide Severity Scale
- Healthcare utilisation
    - Referral to any secondary (hospital) care
    - Referrals to secondary (psychiatric) care
    - Hospital admissions
- Procedure occurrence of electrotherapy, including transcranial direct current stimulation (tDCS), electro compulsive therapy (ECT), repetitive transcranial magnetic stimulation (rTMS), or vagus nerve stimulation (VNS)
- Procedure occurrence of general psychotherapy, including cognitive behavioural therapy (CBT) or interpersonal therapy (IPT)
- Remission of depression

The preliminary concept sets used for the identification of outcomes are described in **Annex IV**. These concept sets are defined using Systematized Nomenclature of Medicine (SNOMED) codes. These codes will be refined during the study execution following the DARWIN EU® phenotyping standard processes, which involves the review of code lists by clinical experts, and the review of phenotypes after their execution in the participating data sources.

Objective 2

For objective 2 there are no outcomes.

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

Objective 1

Prevalence and incidence will be stratified by the following covariates:

- Sex
- Age (at index date) groups: 10–19, 20–39, 40–59, 60–79, and 80+

Objective 2

For objective 2a, the frequency of recording of the following variables will be described:

- Composite suicide-related events (suicide, suicide attempt, suicide ideation, self-harm)

- Observation occurrence of suicide attempt
- Observation occurrence of suicidal ideation
- Observation occurrence of self-harm
- Condition occurrence of depression
- Observation occurrence of healthcare admission or referral for secondary care
- Measurement occurrence of Patient Health Questionnaire-9 (PHQ9) depression scale, Beck Depression Inventory (I and II), Hamilton Depression Scale, Columbia Suicide Severity Scale

For objective 2b, the frequency of the following variables will be described:

- Observation occurrence of composite suicide-related events (suicide, suicide ideation, suicide attempt, and self-harm)
- Observation occurrence of suicide attempt
- Observation occurrence of suicide ideation
- Observation occurrence of self-harm

Objective 2b will be stratified by sex and age (at index date) groups: 10–19, 20–39, 40–59, 60–79, and 80+

For objective 2c, the values if available of the following variables will be described:

- Measurement occurrence of Patient Health Questionnaire-9 (PHQ9) depression scale
- Measurement occurrence of Beck Depression Inventory (I and II)
- Measurement occurrence of Hamilton Depression Scale
- Measurement occurrence of Columbia Suicide Severity Scale

<u>For objective 3:</u>

The frequency of recording (N and %) of the following variables will be described:

Demographic and socioeconomic characteristics:

- Sex
- Age (in years) at index date
- Socioeconomic indicator (if available across data sources)

Prescription or dispensing record of the following drugs used systemically (assessment window [-inf, 0]):

- Antidepressants
- Antipsychotics
- Benzodiazepines
- Stimulants
- Hypnotics

Occurrence of the following conditions (assessment window [-inf, 0]):

- Schizophrenia
- Anxiety disorder
- Substance use disorder (excluding tobacco and alcohol use)
- Alcohol abuse

- Personality disorders
- Bipolar disorder
- Eating disorder
- Pain
- Chronic pain
- Cancer (including haematological, but excluding non-melanoma skin cancers)

Occurrence of the following conditions (assessment window [-inf, -1]):

- Composite suicide-related events to understand if the index date event is the first event in history

Occurrence of the following procedures (assessment window [-inf, -1], [0, inf]):

- Psychotherapy (including CBT and IPT)
- Electrotherapy (including tDCS, ECT, rTMS, and VNS)

Occurrence of the following measurements (assessment window [-inf, -1], [0, inf]):

- Patient Health Questionnaire-9 (PHQ9) depression scale
- Beck Depression Inventory (I and II)
- Hamilton Depression Scale
- Columbia Suicide Severity Scale

Measurement values if available (assessment window [-180, 180]):

- Patient Health Questionnaire-9 (PHQ9) depression scale
- Beck Depression Inventory (I and II)
- Hamilton Depression Scale
- Columbia Suicide Severity Scale

Visit/observation of healthcare utilisation (assessment window [-inf,-1], [0, inf]):

- Referral to any secondary (hospital) care
- Referral to secondary (psychiatric) care
- Hospital admissions

Occurrence or observation of remission of depression (assessment window [0, inf])

Observation occurrence of lifestyle factors (assessment window [-inf, -1]):

- Smoking
- Obesity

The preliminary concept sets used for the identification of covariates are described in **Annex IV**. These codes will be refined during the study execution following the DARWIN EU® phenotyping standard processes, which involve the review of code lists by clinical experts and the review of phenotypes after their execution in the participating data sources.

9.7. Study size

No sample size has been calculated, as this is a descriptive study which will not test a specific hypothesis. In addition, we will use already collected available data to estimate prevalence of suicide, suicide-related events, depression, and associated variables.

9.8. Analysis

9.8.1. 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.

9.8.2. 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.

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

Objective 1

The prevalence estimation will be calculated based on OMOP CDM mapped data using the *IncidencePrevalence* R package, developed by DARWIN EU®.[4] Overall incidence rates (per 100,000 person-years) and period prevalence estimates (per 1000 persons) will be calculated, as well as stratified by age group, sex, and calendar year.

Individuals enter the denominator population at the start of each calendar year from the start of the study period onwards or once the eligibility criteria are fulfilled. Those study participants who enter the denominator population will then contribute follow-up time during each calendar year. Follow-up of subjects who die or become lost to follow-up will be censored at the time of death or loss to follow-up. Subjects with data until the end of the study period without a record of death or loss to follow-up will be administratively censored at the end of the study period.

Period prevalence will be calculated by counting the number of individuals per calendar year with record of the outcomes of interest (see **Section 8.6.2. Outcome - Objective 1**). These counts will be divided by the denominator (the number of persons contributing time-at-risk in the period) to calculate a proportion. Period prevalence will be reported as a percentage with 95% confidence intervals, as estimated by the Wilson Score method.

Descriptive statistics of the study population will be calculated: total number of people, number of people per sex (N and percentage of the total study population), number of people per age group (N and percentage of the total study population).

Objective 2

For objective 2a, we will estimate the median number (and IQR) of suicide-related events in individuals with at least one suicide-related event. The analysis will be calculated based on OMOP CDM mapped data using the *CohortCharacteristics* R package, developed by DARWIN EU®.[5] The package will use the functions *tableIntersectCount = list()* and *addCohortIntersectCount()*.

For objective 2b, we will estimate the mean cumulative function (MCF) of composite suicide-related events, suicide attempt, suicide ideation, self-harm, and depression occurrences. The MCF at time t will report the cumulative number of events per person on average up to time t. Plotting the MCF results in a curve showing how event recording accumulation evolves over time. To execute this, a table with patient ID, time (since index date in days), event indicator (event = 1 or censor = 0), and sex will be used. The MCF will be estimated using the function from the *reda* R package.[6] The parameters of the *mcf* function, the variance and estimator type, will be explored during study execution. The analysis will also be stratified by sex and age group at index date. The output (overall, by sex, and by age group) will be plotted using *ggplot*, as seen in **Figures 12–15**.

For objective 2c, we will describe the values of depression/suicide-related event symptom measurements (PHQ9 depression scales). We will estimate summary values including median, IQR, and min–max values. For this we will use *MeasurementDiagnostics*.[7]

Objective 3

Characterisation of individuals with incident suicide (completed), suicide attempt, suicidal ideation, self-harm, or depression will be conducted using the *CohortCharacteristics* R package developed by DARWIN EU®.[5] In total, five cohorts will be characterised

For each patient characteristic listed in **Section 8.6.4**, the number and proportion (%) of individuals with a record will be presented. For characterisation by conditions and drugs, all diagnoses or prescriptions/dispensing will be considered, irrespective of whether they are incident or prevalent. Sex, median age (plus IQR), proportion of composite suicide-related events before start of study period will be measured at index date (i.e., date of incident suicide, suicide attempt, suicidal ideation, self-harm, or

depression). If data is available, we will describe the values of depression/suicide-related event symptom measurements (PHQ9 depression scales). We will estimate summary values including median, IQR, and min–max values. For this we will use *MeasurementDiagnostics*.[7]

For drugs, comorbidities, procedures, measurements, and lifestyle factors, the number and proportion of individuals with a record within each specified time window will be presented.

9.8.4. Output

Output will include a report with an executive summary, and the following tables and figures. An interactive dashboard will be generated by incorporating all the results (tables and figures) included in the report.

Table 2. Distribution of study participants' characteristics of those with suicide-related events (N, %, median, and IQR).

Characteristic	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5	Data source 6
Overall, N						
Median age (IQR) at index date						
Mean age (SD) at index date						
Age groups in year, N (%), 10–19						
Age groups in year, N (%), 20–39						
Age groups in year, N (%), 40–59						
Age groups in year, N (%), 60–79						
Age groups in year, N (%), 80+						
Median index year (IQR)						
Male, N (%)						
Female, N (%)						

Acronyms: IQR = inter-quartile range, SD = standard deviations

Table 3. Distribution of study participants' characteristics of those with depression (N, %, median, and IQR).

Characteristic	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5	Data source 6
Overall, N						
Median age (IQR) at index date						
Mean age (SD) at index date						
Age groups in year, N (%), 10–19						
Age groups in year, N (%), 20–39						

Characteristic	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5	Data source 6
Age groups in year, N (%), 40–59						
Age groups in year, N (%), 60–79						
Age groups in year, N (%), 80+						
Median index year (IQR)						
Male, N (%)						
Female, N (%)						

Acronyms: IQR = inter-quartile range, SD = standard deviations

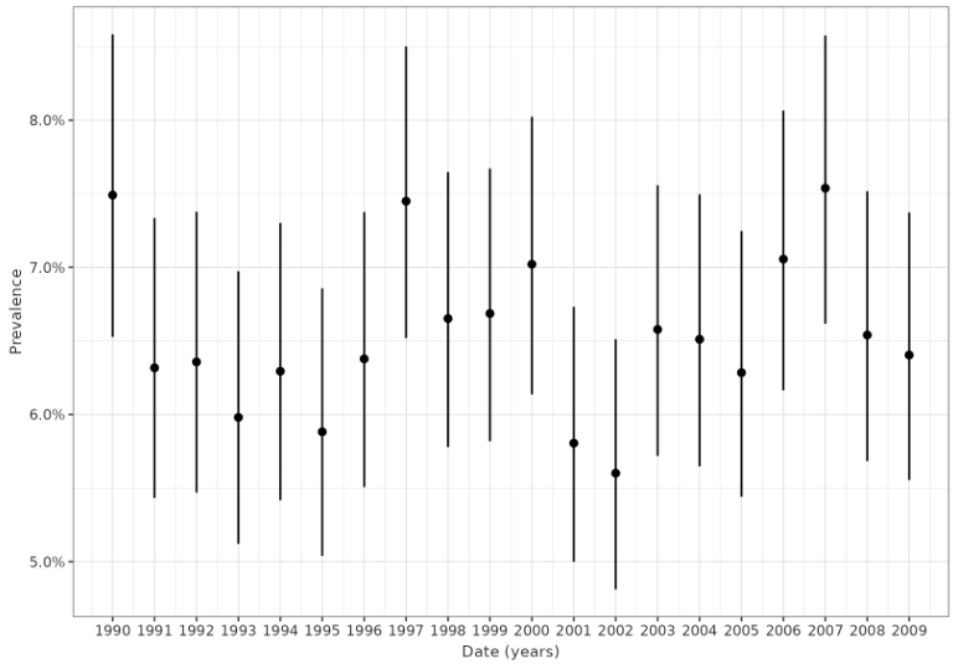

Figure 4. Period prevalence (%) for each outcome specified in Section 8.6.2 for each calendar year.

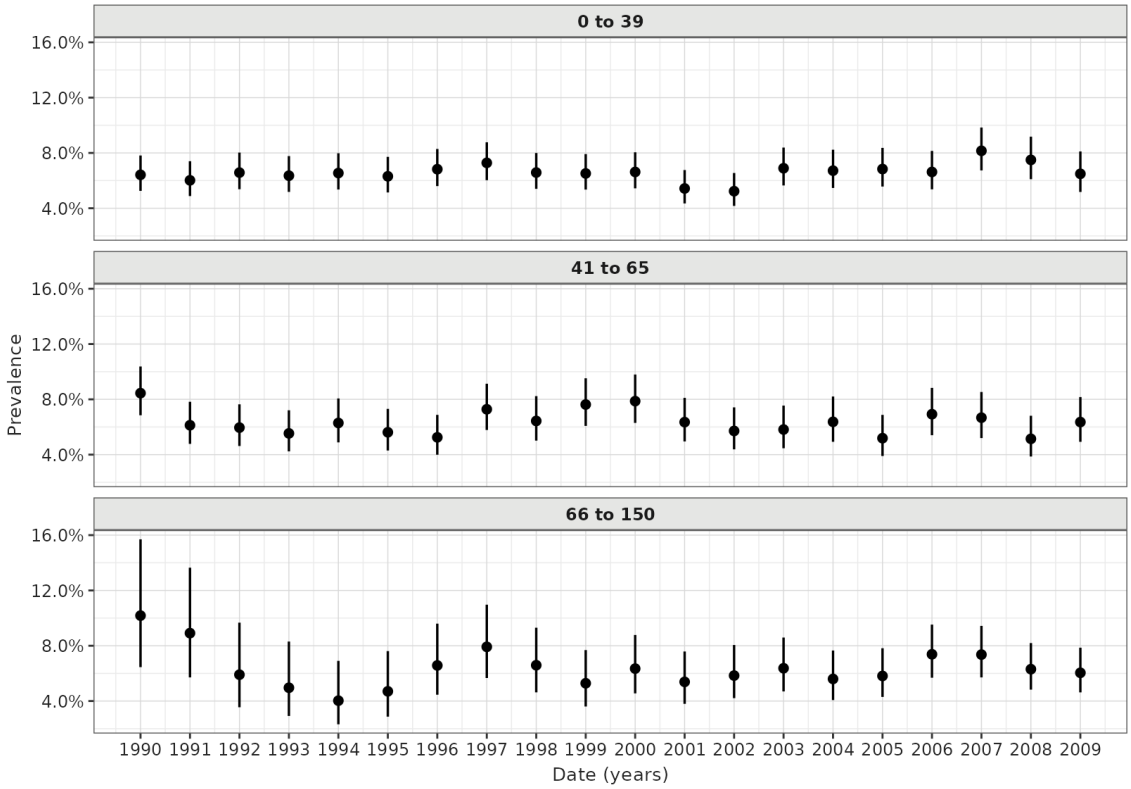

Figure 5. Period prevalence (%) for each outcome specified in Section 8.6.2 for each calendar year, stratified by age. Note: similar figures will be presented stratified by sex.

Table 4. Overall rate of suicide-related events in individuals with suicide-related event or depression after study entry.

Data source	Population	N individuals	Total suicide-related event occurrences	Total follow-up months	Mean rate (95% CI)	Median rate (Q10, Q25, Q75, Q90)
Data source 1	Suicide-related event					
	Suicide attempt					
	Suicide ideation					
	Intentional self-harm					
	Depression					
Data source 2	Suicide-related event					
	Depression					
Data source x	Suicide-related event					
	Depression					

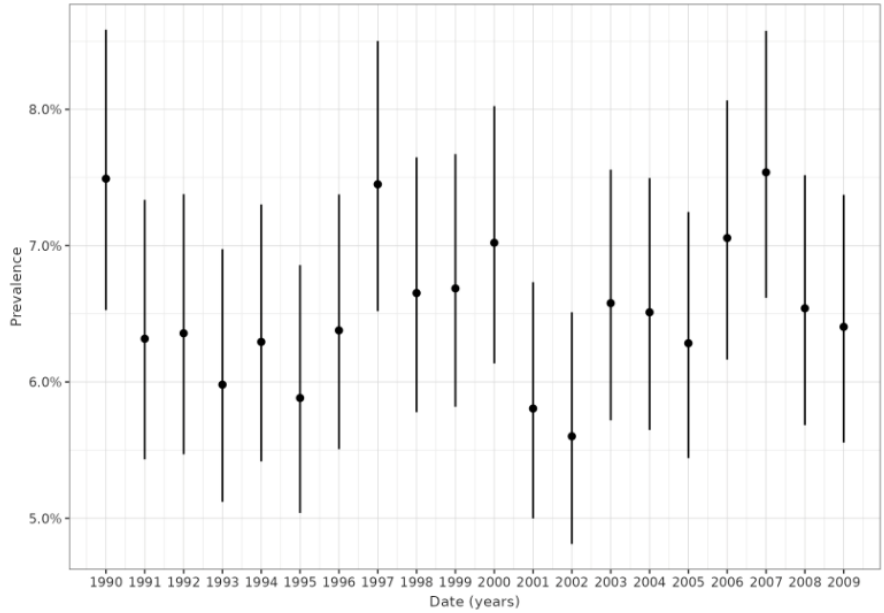

Figure 6. Incidence rate (%) for suicide-related events for each calendar year.

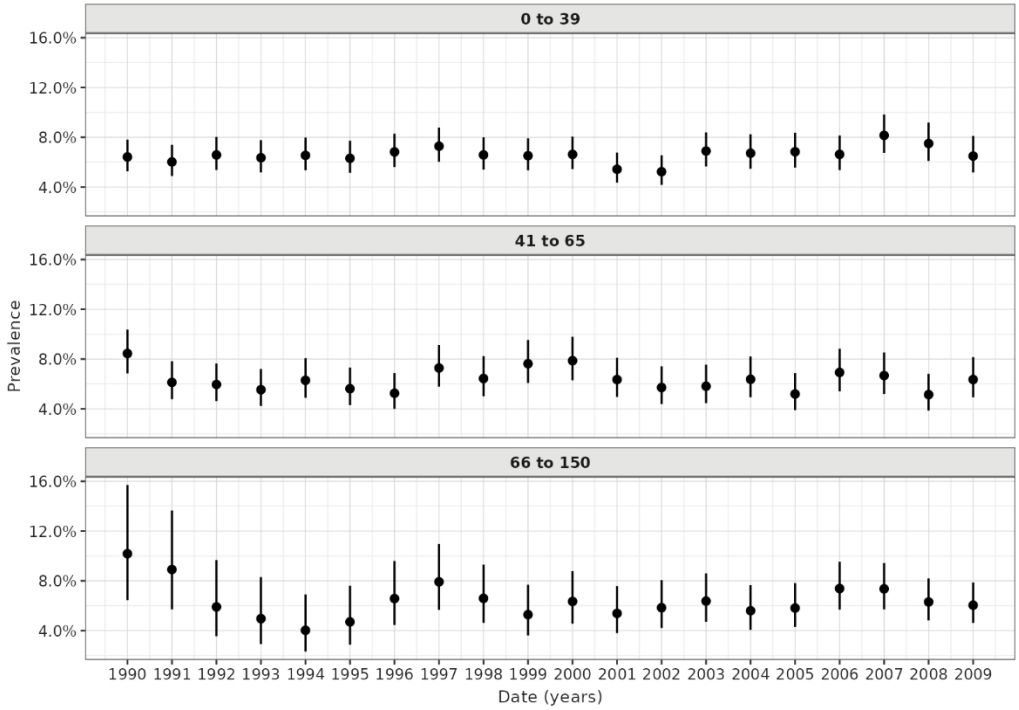

Figure 7. Incidence rates (%) for suicide-related events for each calendar year, stratified by age. Note: similar figures will be presented stratified by sex.

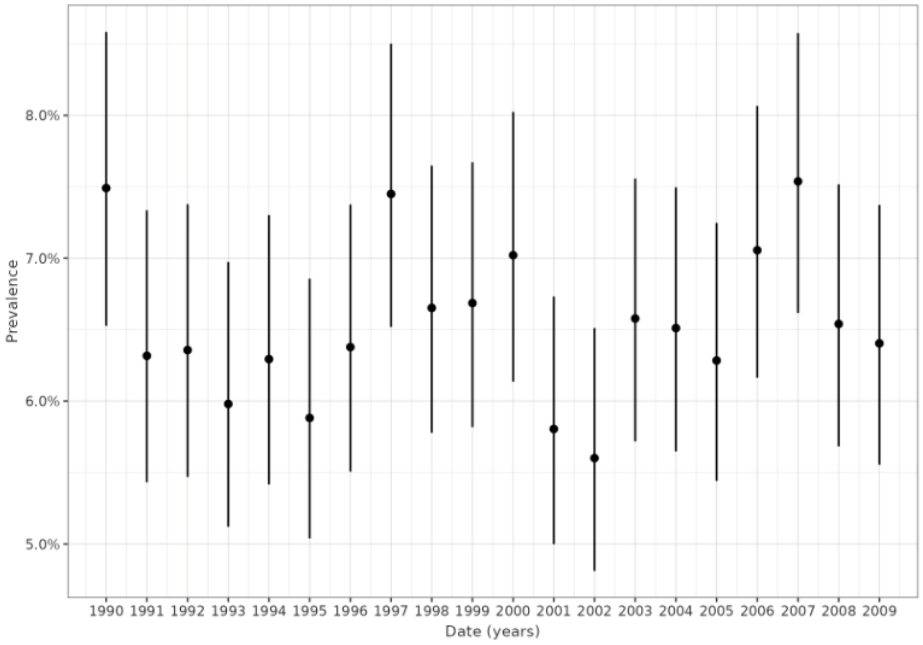

Figure 8. Incidence rate (%) for completed suicide for each calendar year.

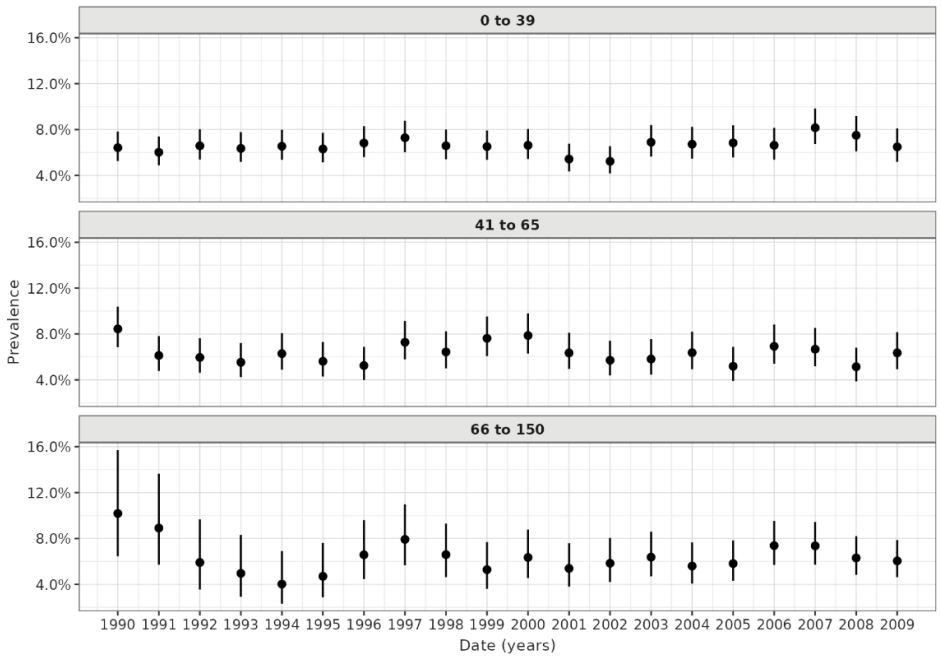

Figure 9. Incidence rates (%) for completed suicide for each calendar year, stratified by age. Note: similar figures will be presented stratified by sex.

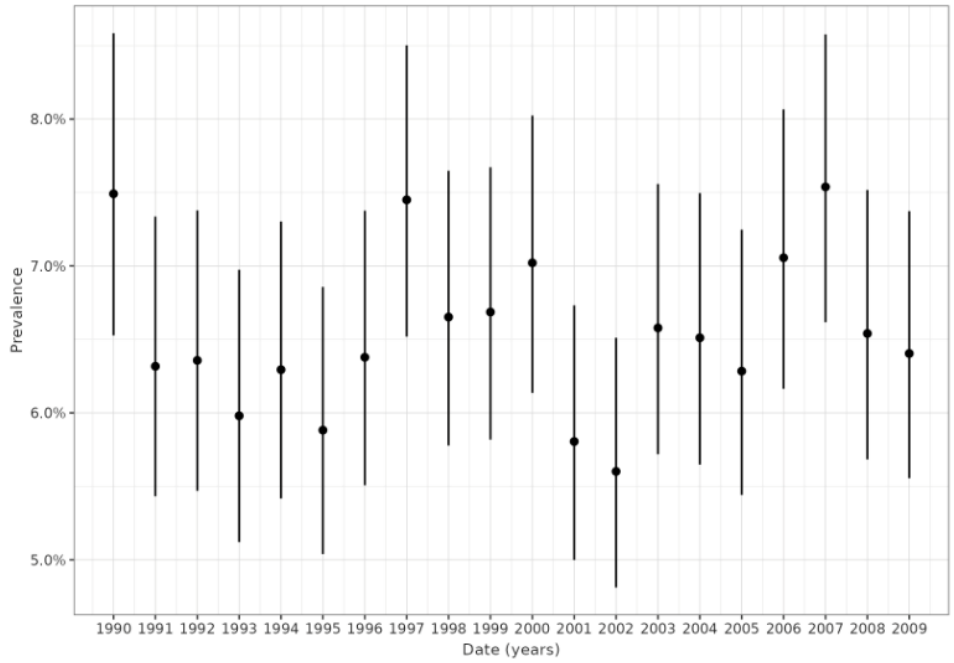

Figure 10. Incidence rate (%) for depression for each calendar year.

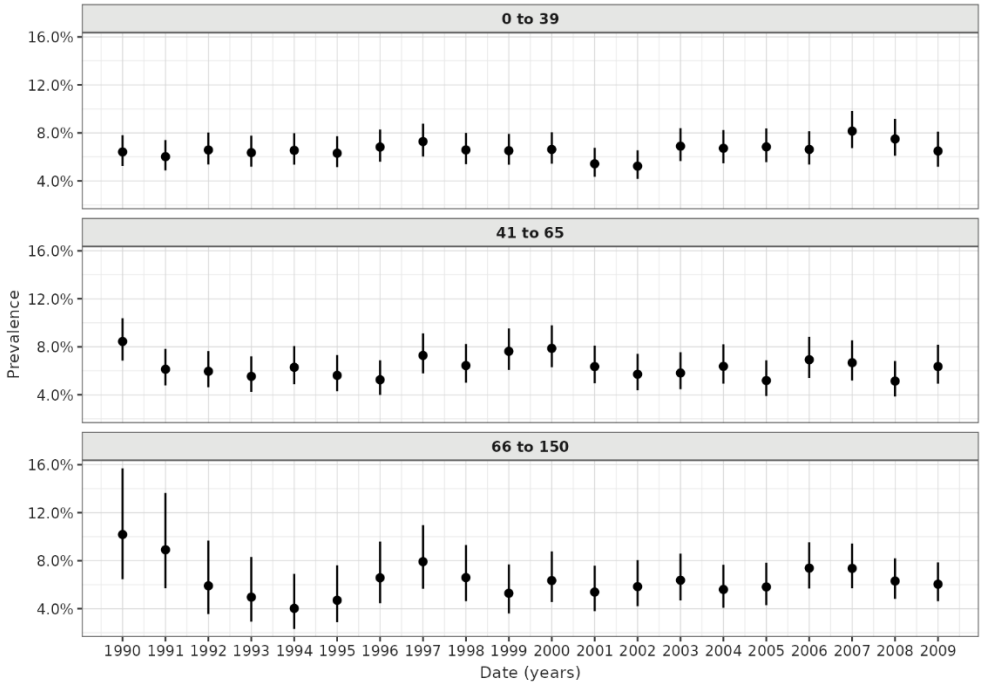

Figure 11. Incidence rates (%) for depression for each calendar year, stratified by age. Note: similar figures will be presented stratified by sex.

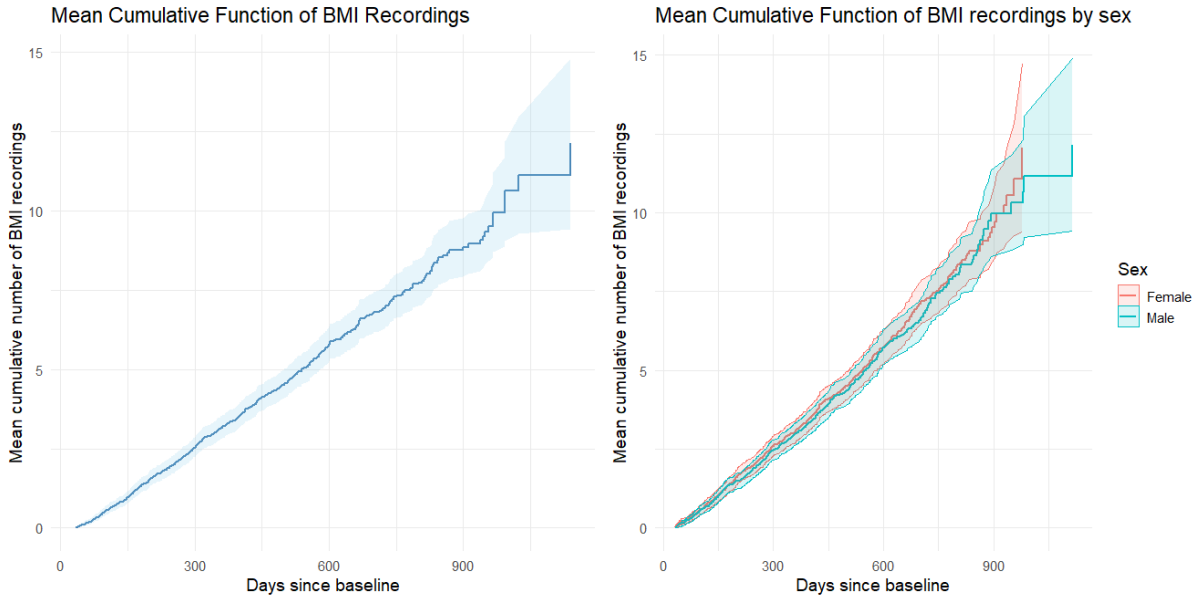

Figure 12. Mean cumulative function of composite suicide-related events in each data source. Overall (left) and stratified by sex (right).

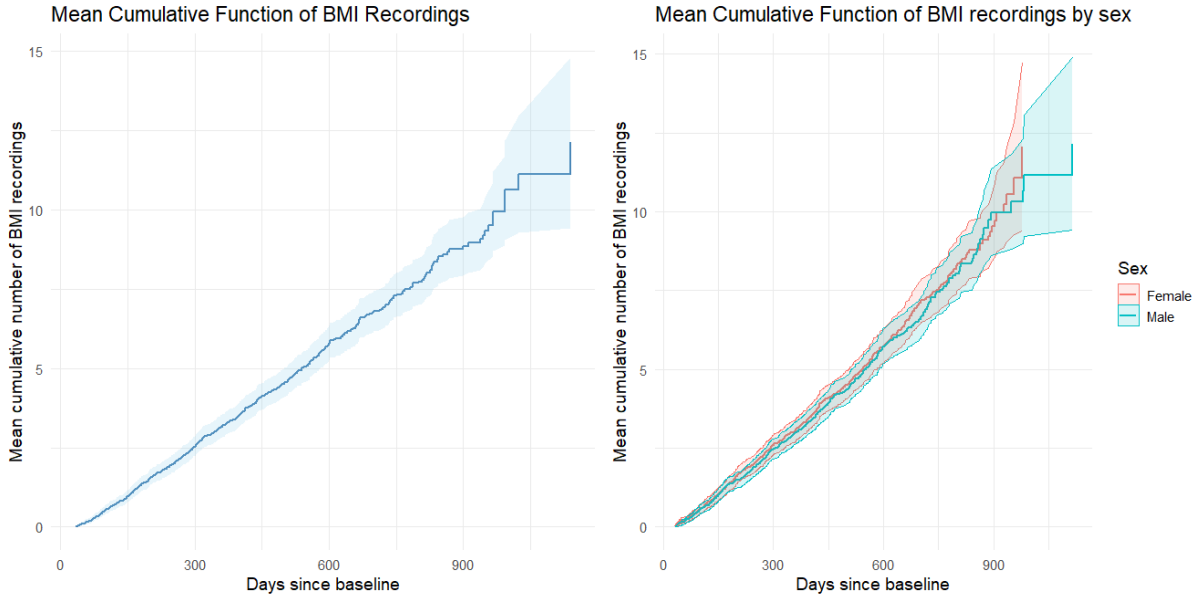

Figure 13. Mean cumulative function of suicide attempt in each data source. Overall (left) and stratified by sex (right).

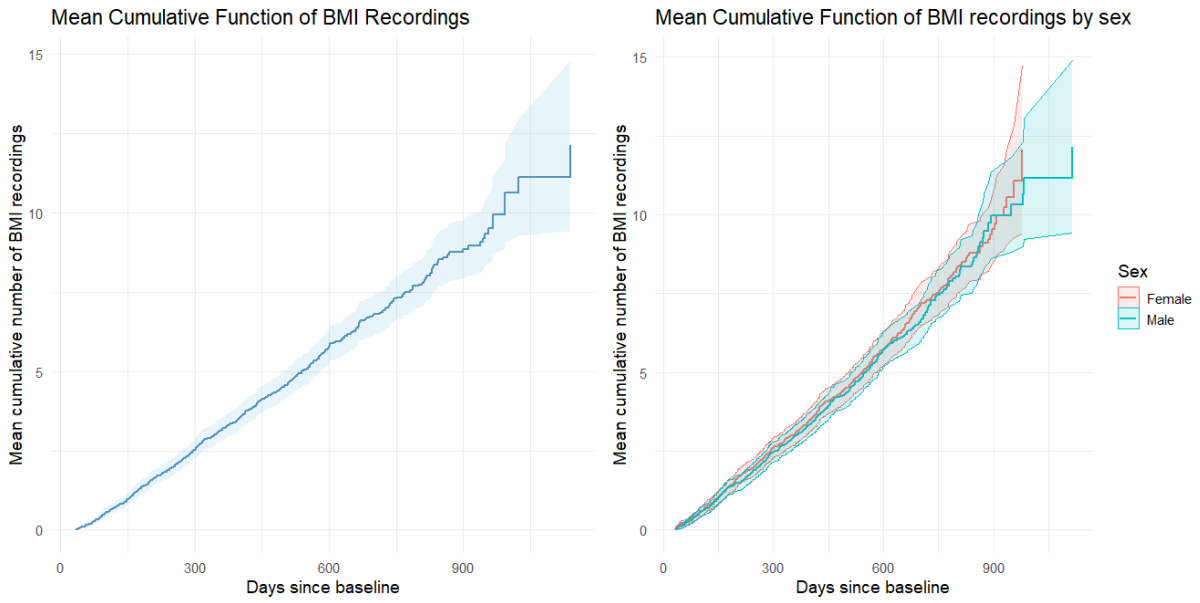

Figure 14. Mean cumulative function of suicide ideation in each data source. Overall (left) and stratified by sex (right).

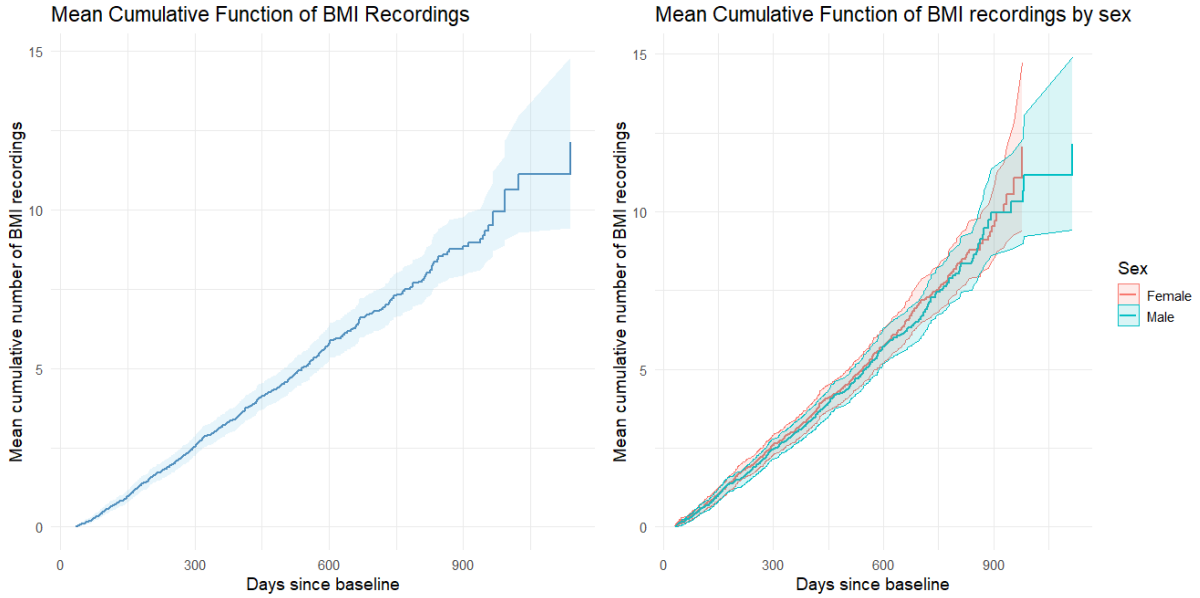

Figure 15. Mean cumulative function of self-harm in each data source. Overall (left) and stratified by sex (right).

Table 5. Demographic characteristics of the study population on index date within individuals with a suicide-related event, completed suicide, or depression.

	Data source 1			Data source 2			Data source 3			Data source 4			Data source 5			Data source x		
Characteristic	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression
Overall, N																		
Median age (IQR) at index date																		
Mean age (SD) at index date																		
Male, N (%)																		
Female, N (%)																		

Acronyms: IQR = inter-quartile range, SD = standard deviations

Table 6. Prespecified characteristics of the study population on index date within individuals with a suicide-related event, completed suicide, or depression.

	Data source 1			Data source 2			Data source 3			Data source 4			Data source 5			Data source x		
Characteristic	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression	Suicide-related event	Completed suicide	Depression
Number subjects, N																		
Antidepressants																		
Antipsychotics																		
Benzodiazepines																		
Stimulants																		
Hypnotics																		
Schizophrenia																		
Anxiety disorder																		
Substance use disorder																		
Personality disorders																		
Psychotherapy (including																		

	Data source 1			Data source 2			Data source 3			Data source 4			Data source 5			Data source x		
CBT and IPT)																		
Electrotherapy (including tDCS, ECT, rTMS, and VNS)																		
Smoking																		
Obesity																		

Acronyms: CBT = cognitive behavioural therapy, IPT = interpersonal therapy, tDCS = transcranial direct current stimulation, ECT = electro compulsive therapy, rTMS = repetitive transcranial magnetic stimulation, VNS = vagus nerve stimulation

9.9. Evidence synthesis

Results from analyses described in **Section 8.8.3.** will be presented separately for each data source. No meta-analysis of results will be conducted. The results of P4-C1-022, P4-C2-015, and P4-C2-016, will be reported together in one final report. Once all three studies yield results, tables and figures will be made according to data source type: primary care data source, secondary care data source, or data sources encompassing information from both.

10. STRENGTHS AND LIMITATIONS

This study, in combination with the two routine-repeated studies (P4-C2-015 and P4-C2-016), constitutes the majority of the DARWIN EU® data partner network. This allows for broad analysis of suicide and suicide-related events as an outcome in a large number of RWD sources across Europe. This will provide insights into the relevance of data sources for future studies with a focus on this outcome, as well as a better understanding of the limitations of the underlying data and where attention should be paid for increasing data curation, data source linkages, use of free-text, and standardisation (mapping).

This study will not comprehensively address all five of the data quality (DQ) dimensions, as outlined in the DQ Framework: reliability, relevance, timeliness, coherence, and extensiveness.[8] It would not be feasible to validate each component of these dimensions, some of which are beyond the scope of this study. Instead, this study considers extensiveness and relevance as the DQ dimensions that are to be most explored.

This study will describe the data relating to suicide and depression, but it is limited in that it cannot validate these events with the source data. Suicide and suicide-related events are often subject to misclassification around the event, as well as the timing of the event in relation to the death date. Suicide is often reported through various databanks into the data source. One pertinent data source are national and regional death registers, which are often not routinely linked to electronic healthcare data sources except on a case-by-case basis. While this study will characterise the available data across the DARWIN EU® network, there will still be substantial underestimation of suicide due to this missing linkage.

11. REFERENCES

1.Amini M, Verhamme, K. DARWIN EU® - Incidence of suicidality in patients with specific chronic skin conditions EMA-HMA Catalogue of RWD studies2024 [Available from: https://catalogues.ema.europa.eu/node/4195/administrative-details#darwin-regulatory.

2.Hunt N, van Leeuwen, G., van Kessel, M., Williams, R., Duarte Salles, T., Prieto Alhambra, D., Verhamme, K. DARWIN EU® - Suicidality following exposure to doxycycline HMA-EMA Catalogue: EMA; 2024 [Available from: https://catalogues.ema.europa.eu/node/4181/administrative-details.

3.WHO. Adolescent health 2025 [Available from: https://www.who.int/health-topics/adolescent-health#tab=tab_1.

4.Raventós B, Català M, Du M, Guo Y, Black A, Inberg G, et al. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. Pharmacoepidemiology and drug safety. 2024;33(1).

5.Catala M GY, Lopez-Guell K, Burn E, Mercade-Besora N, Alcalde M CohortCharacteristics: Summarise and Visualise Characteristics of Patients in the OMOP CDM. R package version 0.4.0 2024 [Available from: https://darwin-eu.github.io/CohortCharacteristics/.

6.Wang W FH, Yan J. reda: Recurrent Event Data Analysis. R package. 0.5.6 ed2025.

7.Burn E M-BN, Alcalde-Herraiz M. MeasurementDiagnostics: Diagnostics for Lists of Codes Based on Measurements. R package version 0.1.0999 ed2025.

8.Data Analytics and Methods Task Force. Data Quality Framework for EU medicines regulation. EMA: EMA; 2023.

12. ANNEXES

ANNEX I. Description of data sources

P4-C1-022:

Croatia: Croatian National Public Health Information System (NAJS)

#	Section	Description
1	Database Identification and country	NAJS (Croatian National Public Health Information System) Croatia
2	Data partner information section	Croatian Institute of Public Health Department of Data Science and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. Geographic coverage covers whole Croatia, with various levels of resolution for different registries. Current estimates for the population in Croatia will be available at: https://podaci.dzs.hr/hr/podaci/stanovnistvo/procjena-stanovnistva/ for each year.
4	Healthcare setting / type of data	Primary care – gps, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. For both inpatient and outpatient setting diagnoses, medication, procedures, and measurements are captured. Since 2025, family relationships from the birth registry, histopathological data from the cancer registry, and vital signs from health risk assessment have also been added. The year of availability of information depends on the setting: • 2014 for lab tests • 2015 for general practitioners • 2016 for secondary conciliatory care • 2017 for hospital records • 2020 for vaccination records
5	Data collection process	Inpatient hospital billing systems, and Other. Data is entered by clinicians at healthcare contact, then combined by CIPH into the NAJS database.
6	General representativeness	The data is collected from public health records, as the majority of health care in Croatia is public. Personal details are collected to a better extent for insured individuals compared to uninsured patients.
7	Data content /source coding	Medication prescriptions are recorded with ATC codes, and diagnoses with ICD10 codes.
8	Data Harmonization	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. Records from 2017 include insured patients with reliable IDs. Uninsured patients do not have reliable IDs. For example, if a patient changed her status from insured to uninsured, or vice versa, she could be counted several times, as could tracking records from before 2017 and after. By using the unique personal identifier for Croatian citizens, it can be checked and verified.
9	Quality control (database specific)	There is a network of registry personnel (leaders, administrators, coders, sources) working on data coverage and other quality dimensions. An analytical team routinely checks for erroneous entries in hospital records, removing double entries, false dates, and overlapping stays. Entries without enough data or with obviously erroneous dates from primary care analysis are being excluded.
10	Linkage	The national death registry is updated monthly, and primary care is updated weekly. Specific registries are included in NAJS (e.g., diabetes registry), where inclusion criteria vary across these registries.
11	Vital status	NAJS is linked to the national death registry.

#	Section	Description
12	Limitations	Hospital data is available from 2017 onwards. This is often used as start of data collection, while laboratory and GP data is captured before that (since 2014 and 2015 respectively). The total and active person count in the NAJS data is larger than the current population of Croatia. This explained by a) the person table included deceased and all previously insured people and b) there is no information about insurance ending. It is known that a lot of people migrated (300k-400k) and weren't included in the last population census but still are in the NAJS database. In-hospital administrations are managed via paper drug charts and hospital discharge summaries are currently not captured into NAJS.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111155 Website: https://www.hzjz.hr/nacionalni-javnozdravstveni-informacijski-sustav-najs/

Denmark: Danish Data Health Registries (DK-DHR)

#	Section	Description
1	Database 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, Devices, and Sociodemographic information.
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, 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 Harmonization	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. No.
9	Quality control (database 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

#	Section	Description
		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.
11	Vital status	The Cause of Death registry (DAR) is used, the cause of death is collected using ICD-10 codes.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
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 database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111217 Website: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/healthdatadenmark

Estonia: Estonian Biobank (EBB)

#	Section	Description
1	Database Identification and country	EBB (Estonian Biobank) Estonia
2	Data partner information section	University of Tartu Institute of Computer Science
3	Coverage and timespan	Data collection since: 2004 Extent: Nation-wide. EBB is a nation-wide database containing records from 2004 onwards. Estonian population-based cohort size of 211,800 participants (01/01/2024) aged 18 years and older recruited at GP offices, private practices, and hospitals or in the recruitment offices of the Estonian Genome Center.
4	Healthcare setting / type of data	Primary care – gps, and community pharmacists, and primary care specialists (e.g., paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Registry which collects electronic records from the biobank and cohort study.
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Biobank, and Other. Data is retrieved from ...
6	General representativeness	The age, sex, and geographical distribution closely reflect those of the Estonian adult population and encompass close to 5% population. Overall, 3.4% of Estonian men and 5.5% of Estonian women are represented in EBB. Older people tend to participate less frequently; however, all age groups are well represented.
7	Data content /source coding	All participants have undergone a standardized health assessment, including provision of blood samples for purification of DNA, white blood cells, and plasma, and completed a questionnaire covering various health-related topics, such as lifestyle, diet, and clinical diagnoses. Diseases and health problems are recorded as ICD-10 codes and prescribed medicine according to the ATC classification.
8	Data Harmonization	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. No, there is one national identifier that allows linking together all encounters across databases.
9	Quality control (database specific)	The quality control procedures in the Estonian Biobank aim to remove the most obvious mistakes in the data, misspellings, impossible dates, duplicates. Before performing the ETL, several problems are fixed on the source data. Since the ETL procedures are used for a number of different datasets (from the same national sources), we have a growing number of pre-processing steps that correspond to the issues we have

#	Section	Description
		discovered previously in the data, such as checking for the presence of critical values, harmonizing date and unit of measurement formats, checking the validity of certain entries against classifiers, etc.
10	Linkage	Follow-up data are available via linkage with national health-related registries and via re-examination of participants. Furthermore, electronic health records are updated for phenotypic outcome information every half year. The EBB database is regularly linked with national registries, hospital databases, and the database of the national health insurance fund, which holds treatment and service bills.
11	Vital status	Vital status (death date and causes of death) are obtained from the Causes of Death Registry.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Leitsalu L, Haller T,Esko T,Tammesoo ML,Alavere H,Snieder H,Perola M,Ng PC,Mägi R,Milani L,Fischer K,Metspalu A "Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu." International journal of epidemiology (2015): 24518927
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111114 Website: https://genomics.ut.ee/en/content/estonian-biobank

Germany: InGef Research Database (InGef RDB)

#	Section	Description
1	Database Identification and country	InGef RDB (InGef Research Database) Germany
2	Data partner information section	Institut für angewandte Gesundheitsforschung Berlin GmbH
3	Coverage and timespan	Data collection since: 2014 Extent: Nation-wide. The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDC currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs. The data in the database depicts all health care use which has been reimbursed by the SHI.
4	Healthcare setting / type of data	Primary care – gps, and community pharmacists, and primary care specialists (e.g., paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data. The following data elements are collected: pregnancy data, hospital admission, and/or discharge also with ICU admission. Prescription, dispensing drugs, and Advanced therapy medicinal products. Contraception, medical devices, vaccinations, procedures, diagnoses, and demographic information.
5	Data collection process	Insurance/administrative claims. The data in the database depicts all health care use which has been reimbursed by the SHI (statutory health insurances).
6	General representativeness	The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDC currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs.
7	Data content /source coding	The ATC and OPS (Operationen- und Prozedurenschlüssel) are used for prescription and dispensing drugs. For Procedures, the EBM (Einheitlicher Bewertungsmaßstab - doctor's fee scale) and for ambulatory procedures; OPS (Operationen- und Prozedurenschlüssel) for operations conducted at the hospital are used. Medical events are coded in ICD-10-GM and another vocabulary used is PZN (Pharmazentralzummer -pharmaceutical reference number).

#	Section	Description
8	Data Harmonization	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. No. In the German statutory health system, a person can only be enrolled in one health insurance at a time. However, if a person changes from one contributing insurer to another, a new ID number will be generated.
9	Quality control (database specific)	Before entering the InGef database, the data elements are checked with respect to data format, completeness, and plausibility. After each data update, data are compared with the previous data update in regard to number of records, number of data providers, etc. Due to the anonymized nature of the database, no direct validation of the data (e.g., using medical charts as the gold standard) is possible. Data delivery by health care providers is generally based upon standardized data requirements and formats provided by the National Association of Statutory Health Insurance Funds (compare: https://www.gkv-datenaustausch.de/leistungserbringer/leistungserbringer.jsp)
10	Linkage	No
11	Vital status	The Cause of Death is not captured; the date of death is captured.
12	Limitations	Ambulatory diagnoses are received from the source on a quarterly basis. These diagnoses are mapped to the observation table with the concept_id History of event within 3 months (1340222), with the actual diagnosis concept_id recorded in the field value_as_concept_id, and the date as the last day of the respective quarter (i.e. 30/31st of Mar/Jun/Sept/Dec). Ambulatory prescriptions are available with exact dates. The cause of death is not captured and there is no linkage with other data sources. Approx. 10.5 Million insurees are included in the database, 7.8 Million of these actively insured in 2024. This corresponds to 7% of the total German population. Data are longitudinally linked over a period of currently ten years.
13	Main references	Andersohn F, Walker J "Characteristics and external validity of the German Health Risk Institute (HRI) Database." Pharmacoepidemiology and drug safety (2016): 26530279
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111207 Website: https://www.ingef.de/en/

The Netherlands: Integrated Primary Care Information (IPCI)

#	Section	Description
1	Database Identification and country	IPCI (Integrated Primary Care Information) The Netherlands
2	Data partner information section	Erasmus University Medical Center Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards. However, it mainly covers the central part of the country, including the most densely populated area (the 'Randstad') and non-urban areas. IPCI contains information on all patients registered with GPs responsible for non-emergency care and referrals. A patient is registered at birth or at first encounter with the GP.
4	Healthcare setting / type of data	Primary care – gps. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. Data is aggregated by Erasmus MC data managers and combined in one harmonized database. Several checks are done

#	Section	Description
		on this database to ensure correct data processing. Persons are mostly uniquely identified, with the exception of when persons change GP practice (when the same individual can receive several different identifiers).
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content /source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data Harmonization	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 can be registered under different IDs. However, in the Netherlands, patients typically have one GP and changing practice is uncommon.
9	Quality control (database specific)	Prior to each data release, extensive quality control steps are performed, e.g., comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g., number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical data, e.g., distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.
13	Main references	de Ridder MAJ, de Wilde M,de Ben C,Leyba AR,Mosseveld BMT,Verhamme KMC,van der Lei J,Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182143
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

#	Section	Description
1	Database Identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) The United Kingdom

#	Section	Description
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 – gps, 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 is entered by clinicians into the EHR. Data is processed by CPRD and provides data releases for research.
6	General representativeness	CPRD GOLD has been assessed and found to be broadly representative of the UK general population in terms of age, gender, and ethnicity. In CPRD GOLD in January 2025 there were 2,730,707 current acceptable patients (i.e. registered at currently contributing practices that use Vision software, excluding transferred out, deceased patients, and those flagged by CPRD as not acceptable for clinical research for data quality issues). This equals to 4.07%, based on the UK population estimates of 67,026,300 from the Office of National Statistics (mid-2023). Current patients are only from Scotland, Wales, and Northern Ireland. Historically, GOLD does contain data from England as well.
7	Data content /source coding	Gemscript, Read, dm+d
8	Data Harmonization	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 (database 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 check data quality standards at the patient level and associate each patient to a flag, reporting if its data is 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 Practice Level
11	Vital status	Vital status is retrieved from the GP records. Population registry (ONS) data can be requested on a study-by-study basis and linked. This data only covers England and is planned to be mapped to OMOP in the future. The cause of death is not captured.
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.

#	Section	Description
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 database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://cprd.com

P4-C2-015:

Greece: Papageorgiou General Hospital (PGH)

#	Section	Description
1	Database Identification and country	PGH (Papageorgiou General Hospital) Central Macedonia, Greece
2	Data partner information section	Papageorgiou General Hospital Program Management Office
3	Coverage and timespan	Data collection since: 1999 Extent: Other. Patients from across the Macedonia region and from neighbouring countries as well.
4	Healthcare setting / type of data	Primary care – gps, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify). long-term/skilled nursing facility
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems. The data is gathered live (as the patient is examined) in the production EHR instance.
6	General representativeness	PGH, situated in Thessaloniki—the second-largest city in Greece—began operations in 1999. With a capacity of 745 beds, PGH supports over 200,000 hospitalization days annually. The OMOP CDM has around 1.41M patients.
7	Data content /source coding	Medications are coded using ATC codes. KEN («Κλειστά Ενοποιημένα Νοσήλεια») Coding System and Internal Hospital Coding System: This coding system is applied to procedural data, and it was enforced by the Greek Public Health Ministry. It has been mapped to ICD-10. Internal Hospital Coding System: This system is employed for coding medical devices. ICD-10 (International Classification of Diseases, Tenth Revision): This system is utilized for coding both 'Diagnoses' and 'Deaths' from the source. Internal LIS Coding System: This coding system, associated with the internal laboratory information system, is used for laboratory tests and examinations.
8	Data Harmonization	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. No, in Greece, each patient is identified by a unique universal health insurance number.
9	Quality control (database specific)	There is no formal quality assurance plan in place for the raw data collected during daily clinical practice, therefore, gaps or errors in the data should be anticipated.
10	Linkage	No known linkages.
11	Vital status	There are only in-hospital deaths available in the data.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/institution/1000000215 Website: https://www.papageorgiou-hospital.gr/?lang=en

Portugal: Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)

#	Section	Description
1	Database Identification and country	EMDB-ULSGE (Egas Moniz Health Alliance database - Gaia E Espinho) Distrito de Aveiro, Portugal
2	Data partner information section	Clinical Academic Center Egas Moniz Health Alliance
3	Coverage and timespan	Data collection since: 2008 Extent: Regional. The ULSGE includes 32 primary care centres, assisted by three hospitals (Hospital Eduardo Santos Silva, Hospital Distrital Vila Nova de Gaia, Hospital Nossa Senhora da Ajuda) and one specialized rehabilitation centre (Centro de Reabilitação do Norte). It serves the population of the municipalities of Gaia and Espinho, which amounts to approximately 350,000 patients. The total population of the region is estimated to be 335,000.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care). The database captures information about demographics, visit occurrences, diagnoses, medications, and procedures.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. The data is extracted directly from the source EHR system.
6	General representativeness	The hospitals collecting the data are part of the national health service and treat patients of all socioeconomic levels. Therefore, the patient sample should be representative of the population.
7	Data content /source coding	Source data terminologies include ATC, RxNorm, ICD-9, and ICD-10 codes.
8	Data Harmonization	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. The same patients cannot be registered under a different id in the OMOP CDM data.
9	Quality control (database specific)	We have a two-step approach for data quality: A set of unit tests are implemented as data quality scenarios. These scenarios are tested at the ETL level and ran every time the ETL is executed. Those tests ensure that known issues in data quality are addressed and check automatically on every run. In addition, data quality is assessed through the OHDSI DataQualityDashboard project and reported as well. There is room for implementation of specific data quality checks that may be relevant at the study level, if required.
10	Linkage	No known linkages.
11	Vital status	Source for vital status unknown.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111133 Website: https://www.ua.pt/pt/cacemha

Portugal: Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA)

#	Section	Description
1	Database Identification and country	EMDB-ULSRA (Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro)) Distrito de Aveiro, Portugal
2	Data partner information section	Clinical Academic Center Egas Moniz Health Alliance
3	Coverage and timespan	Data collection since: 2008 Extent: Regional. The ULSRA includes 41 primary care centres, assisted by four hospitals (Hospital Dr. Francisco Zagalo, Hospital Visconde de Salreu, Hospital Distrital de Águeda, Hospital Infante D. Pedro). It serves approximately 390,000 patients from most of the municipalities of Aveiro. The total population of the region is estimated to be 713,500.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care). The database captures information about demographics, visit occurrences, diagnoses, medications, and procedures.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. The data is extracted directly from the source EHR system.
6	General representativeness	The hospitals collecting the data are part of the national health service and treat patients of all socioeconomic levels.
7	Data content /source coding	Source data terminologies include ATC, RxNorm, ICD-9, and ICD-10 codes.
8	Data Harmonization	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. The same patients cannot be registered under a different id in the OMOP CDM data.
9	Quality control (database specific)	We have a two-step approach for data quality: A set of unit tests are implemented as data quality scenarios. These scenarios are tested at the ETL level and ran every time the ETL is executed. Those tests ensure that known issues in data quality are addressed and check automatically on every run. In addition, data quality is assessed through the OHDSI DataQualityDashboard project and reported as well. There is room for implementation of specific data quality checks that may be relevant at the study level, if required.
10	Linkage	No known linkages.
11	Vital status	Source for vital status unknown.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111133 Website: https://www.ua.pt/pt/cacemha

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

#	Section	Description
1	Database 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 – gps, 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 Harmonization	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 will 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 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.

#	Section	Description
9	Quality control (database 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 database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/21501 Website: http://www.bifap.org/index_EN.html

Spain: Hospital Universitario 12 de Octubre (H12O)

#	Section	Description
1	Database Identification and country	H12O (Hospital Universitario 12 de Octubre) Comunidad de Madrid, Spain
2	Data partner information section	Fundación Investigación Biomédica Hospital 12 de Octubre Instituto de Investigación (i+12)
3	Coverage and timespan	Data collection since: 2015 Extent: Regional. H12O is a national reference for the treatment of certain pathologies covering patients in the southern area of Madrid region.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify). The H12O database, INFOBANCO, contains information from the different health domains (laboratory, prescriptions, treatments, administrative, diagnoses, etc.). In addition, information is also obtained from other data sources, such as the pathological anatomy system, which provides information about sample analysis, and the cost system, containing information on the cost associated with a contact with the hospital. Work for the inclusion of further data is ongoing, among others, radiological information, or PROMs.

#	Section	Description
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems, and Registries, and Other. Data is entered by clinicians and processed in the regional 'INFOBANCO' platform. This platform allows combining data from multiple heterogeneous sources, and provides mechanisms for data governance, adequacy, interrogation, visualization and analysis for real-world evidence generation and decision support.
6	General representativeness	H12O only covers patients visiting the hospital due to a medical condition and is therefore not necessarily representative of the general population.
7	Data content /source coding	Diagnosis: ICD-10-CM, IDC-9, SNOMED CT, ORPHA; Procedures: ICD-10-PCS, ICD-9, SNOMED CT; Medication: ATC and SNOMED CT; Laboratory tests: LOINC; Histopathology: ICD-O-3.1 and SNOMED CT; Clinical observations: SNOMED CT
8	Data Harmonization	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 cannot be registered with different IDs. Each patient has a unique identifier in the hospital's EHR (NHC), a regional unique identifier (CIPA), and a national unique identifier (CIP-SNS).
9	Quality control (database specific)	The EHR contains certain rules for recording information, e.g., it does not allow closing reports that do not have a coded diagnosis. In addition, it allows the use of terminologies, such as SNOMED or LOINC, for recording information, which helps to reduce unstructured data.
10	Linkage	The hospital data can be linked with pharmacy and the Madrid public health service.
11	Vital status	The hospital is connected to the CIBELES system of the Madrid public health service, which allows the unique identification of users and contains the date of death of the patient. In this way, the hospital has the information of the death, regardless of whether it occurred in the hospital or not.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Miguel Pedrera-Jiménez, Noelia Garcia Barrio, Antonio J Díaz Holgado, Pablo Serrano-Balazote, et al. "INFOBANCO: Standardized platform for management, semantic interoperability, and transparent reuse of EHRs" Conference: EIT Health German-Spanish Symposium on Health DataAt: Mannheim, Germany (2022):
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111145 Website: https://www.comunidad.madrid/hospital/12octubre/

Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)

#	Section	Description
1	Database Identification and country	IMASIS (Institut Municipal Assistència Sanitària Information System) Catalunya, Spain
2	Data partner information section	Consorci Mar Parc de Salut Barcelona (PSMar) Direcció de Control de Gestió (Management and Control Department)
3	Coverage and timespan	Data collection since: 1990 Extent: Regional. The catchment area is essentially the city area of Barcelona. However, any visit recorded from a patient will be included and might include patients residing elsewhere.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Patients are captured when they interact with any service from any hospital or clinic within Parc de Salut Mar Barcelona (PSMar). This can be taken from inpatient EHRs, hospital outpatient specialist care, as well as long term / skilled nursing facilities.

#	Section	Description
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. Long term/skill nursing facility care.
6	General representativeness	The data source contains patients of no specific socioeconomic status, indicating that the dataset should be representative of the population of Barcelona.
7	Data content /source coding	Health outcomes, diagnoses, and procedures are recorded using ICD-9-CM and ICD-10-CM codes. Laboratory measurements are recorded in the source data using LOINC terminology. Additionally, a limited set of demographics are available.
8	Data Harmonization	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 cannot be registered under different ID numbers. They are registered under the same ID, even when changing practice of re-registration or admission.
9	Quality control (database specific)	There are several quality control actions that we have implemented, and the creation and update of IMASIS-2 includes a set of filters that rule out all the data whose format or content are not matching the target variable. Some examples of the actions applied are: • If there are records containing codes with errors that do not belong to the vocabulary used for codifying, the corresponding domain is reviewed or discarded. • If there are records with a date before the birthdate of a patient, this information is also reviewed. • Before the update of IMASIS-2, the number of records in the files coming from the source are counted. Statistics of counts of the new instance of IMASIS-2 are taken and compared with the counts in the source files and with previous IMASIS-2 instances in order to identify issues in the source data. • In the first instances, if the Data Quality Dashboard (DQD) package detected some quality issues, such as conditions diagnosed in patients with an implausible gender in the source data, this information was fixed.
10	Linkage	Information of diagnosis and treatment received in outpatient settings and mortality registries.
11	Vital status	The age at death for a patient is recorded, but without a specific cause.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Mayer MA, Furlong LI,Torre P,Planas I,Cots F,Izquierdo E,Portabella J,Rovira J,Gutierrez-Sacristan A,Sanz F "Reuse of EHRs to Support Clinical Research in a Hospital of Reference." Studies in health technology and informatics (2015): 25991136
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104316 Website: https://www.parcdesalutmar.cat/en/

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

#	Section	Description
1	Database 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. The SIDIAP database contains records of around 6 million people residing in Catalonia, estimated to be representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – gps, and hospital inpatient care. SIDIAP captured data includes routine visits, demographics, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, and lifestyle information.

#	Section	Description
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Catala de la Salut is the owner of the data and acts as 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 Harmonization	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. No.
9	Quality control (database 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.
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 database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

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

#	Section	Description
1	Database Identification and country	HI-SPEED (Health Impact - Swedish Population Evidence Enabling Data-linkage) Sweden
2	Data partner information section	SMPA-GU, Läkemedelsverket, Box 26 Pharmacoepidemiology and Analysis Department (FeA)
3	Coverage and timespan	Data collection since: 2020 Extent: Nation-wide. The catchment area includes the whole of Sweden, covering the full population of approximately 10 million.
4	Healthcare setting / type of data	Primary care – gps, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: Socio-demographics, drug use and prescriptions, diagnoses, cause of death, primary care procedures and visits, as well as secondary care and inpatient visits or clinical events.
5	Data collection process	Registries. The data is acquired from the Swedish nation 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, ICD10 is used for diagnoses, and the Swedish procedure coding system (KVA) is used for clincal procedures.
8	Data Harmonization	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. No.
9	Quality control (database specific)	The source data is obtained from the 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.
11	Vital status	Data on death and cause-of-death are extracted from the Cause-of-Death registry.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: Website:

P4-C2-016:

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

#	Section	Description
1	Database Identification and country	FinOMOP-HUS (Hospital District of Helsinki and Uusimaa) Uusimaa, Finland
2	Data partner information section	HUS – Helsinki University Hospital, Hospital District of Helsinki and Uusimaa HUS Tietohallinto (Information Management)

#	Section	Description
3	Coverage and timespan	Data collection since: 2014 Extent: Regional. HUS is responsible for specialized healthcare in Finland's Uusimaa region and treatment of many rare and severe diseases that are nationally centralized to HUS.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. All visits, examinations, laboratory test, procedures, and treatments are recoded in the HUS IT systems and integrated into the data lake. The data lake stores decades of clinical information in digital format, and data from both past and current source systems are available. Systems providing data into the data lake include: CGI Uranus and Epic Apotti (EHR: visits, diagnoses, medication, etc.), Opera (procedure records), Kemokur and Beacon (cancer-specific medications), Marela (hospital pharmacy), Multilab/Mylab+ (laboratory system), Qpati/Mylab+ (pathology records system).
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. All visits, all procedures, and all given treatments have been recorded systematically in the electronic format. We use more than hundred different operational IT systems.
6	General representativeness	The data reflects patients that have visited the hospital and is thus not a generic population.
7	Data content /source coding	Drugs: ATC, generic name, brand name + dosage and strength information in the data + local route mapping Procedures: Nomesco NCSP Genetic: special mappings to OMOP Genomics Diagnosis: ICD10 + Finnish adaptation “ICD10fi” Laboratory tests: National laboratory test numbers + local HUS variants, National microbe listings, local “additional information” -coding Various observations during care: LOINC and local coding
8	Data Harmonization	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. All patients are identified with a unique, national social security number, which is a permanent person identifier. Based on this, the derived patient identifiers (pseudonyms) in the HUS data lake also are unique.
9	Quality control (database specific)	The HUS data lake and IT management is ISO 13485:2016 and ISO 9001:2015 certified and has capabilities in developing, validating, and CE-marking medical devices and software. Data picks for individual research projects are performed by a dedicated data analyst team and are subject to internal review before data are released for OMOP mapping and into research-specific analysis environments.
10	Linkage	Data from many nation-wide registries can be combined at HUS, which is subject to obtaining special data permits.
11	Vital status	Source for vital status unknown.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111134 Website: https://www.hus.fi/en

Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)

#	Section	Description
1	Database Identification and country	FinOMOP-TaUH Pirha (Tampere University Hospital patient cohort) Pirkanmaa, Finland

#	Section	Description
2	Data partner information section	Pirkanmaa Welfare Services County, Tampere University Hospital Department of Research, Development and Education
3	Coverage and timespan	Data collection since: 2007 Extent: Regional. TaUH Research Database includes all specialities/all patient groups treated in the Tampere University Hospital,
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Secondary, and tertiary care given in the region, including given clinical and pathology diagnoses, diagnostic and therapeutic procedures, laboratory findings, radiology and pathology reports, medication given in the hospital and electronic prescriptions, and continuous medical records (free text), including discharge letters since 2007.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. The data is entered by the clinician at point of contact and processed in a data warehouse for secondary use.
6	General representativeness	The data covers patients visiting the hospital and therefore will not reflect the general population.
7	Data content /source coding	ICD10fi: Finnish modification of ICD10; SNOMED2-TMP: a local SNOMED2-based vocabulary of organ-diagnosis pairs in pathology; NCSPfi: Finnish national version of NCSP/Nomesco vocabulary for procedures; LABfi: Finnish national vocabulary (Kuntaliitto) for laboratory measurements; LABfi_TMP: Local additions to the national vocabulary for laboratory measurements; UNIT_FIN: measurement units; MICROBEfi_TMP: Microbe names in measurement data; ATC: all drugs except anticancer drugs; active ingredient (in Finnish): for anticancer drugs; VNR: Nordic medicine product identifiers (https://pharmaca.fi/en/health/pharma/vnr/); Hilmo_eala: Finnish national medical specialty vocabulary.
8	Data Harmonization	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 some cases, patients can be re-registered with a new ID. Patients are identified by their social security number, and each patient gets a patient ID in the EHR, associated with the social security number. However, there are cases when the social security number has changed, and the associated patient ID also has changed. Approximately 4% of all patient IDs in the database are such duplicates. Majority of such duplicates are newborns who have first been assigned a temporary social security number, which has later been changed to a permanent social security number. An unidentified person can have been given a temporary social security number, and sometimes a person can change their social security number, for example, because of threat (stalking) or gender reassignment. It is technically possible to link all the instances of a person in the database; however, this is not currently done.
9	Quality control (database specific)	The source database tables are checked for completeness on a general level: whether there are null values in the fields, what the date ranges are for each table, are there gaps in the data, and how many unique patients can be found in each data table. Before mapping to the OMOP CDM, we have identified the data tables and the fields in the tables that are required to build the OMOP CDM. The White Rabbit and Rabbit-in-a-Hat tools were used in this process.
10	Linkage	Data from many nation-wide registries can be combined, which is subject to obtaining special data permits.
11	Vital status	Source for vital status unknown.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111135 Website: https://www.pirha.fi/en/web/english

Finland: Finnish Care Register for Health Care (FinOMOP-THL)

#	Section	Description
1	Database Identification and country	FinOMOP-THL (Finnish Care Register for Health Care) Finland
2	Data partner information section	Finnish Institute for Health and Welfare (THL) Department of Knowledge Brokers
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 – gps, and primary care specialists (e.g., paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The THL database covers both public and private, primary, and specialised inpatient 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 register has covered both public outpatient and inpatient specialized care and private inpatient care (TerveysHilmo). Since 2009, the Finnish National Vaccination Register is covered (complete since 2020). The vaccination register covers 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 from 2020 onwards. Since 2011, the register has covered public primary care (AvoHilmo). Since 2020, the register has 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, which is maintained by THL.
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.
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, 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 Harmonization	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 (database 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	No database-specific limitations documented. General limitations for the data type applicable.

#	Section	Description
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):
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111187 Website: https://thl.fi/fi/tilastot-ja-data/ohjeet-tietojen-toimittamiseen/hoitoilmoitusjarjestelma-hilmo

France: Assistance Publique Hôpitaux de Marseille (APHM)

#	Section	Description
1	Database Identification and country	APHM (Assistance publique Hôpitaux de Marseille) France
2	Data partner information section	Assistance Publique – Hôpitaux de Marseille Public Health
3	Coverage and timespan	Data collection since: 2021 Extent: Regional. The data covers all inpatients and outpatients treated at the Assistance Publique – Hôpitaux de Marseille (APHM), which includes five public university hospitals, all located in Marseille, France. The APHM serves not only the local population of Marseille and the surrounding Bouches-du-Rhône department but also attracts patients from the broader Provence-Alpes-Côte d'Azur (PACA) region and Corsica, representing a combined population of over 5 million inhabitants. Additionally, the hospital's specialized services attract patients from across France and abroad, including foreign nationals, all of whom are integrated into the OMOP CDM.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The data source used in this study includes all hospital stays across various care settings—acute care, psychiatric care, rehabilitation care, and home hospitalization. The EHR system covers diagnoses, procedures, drug prescription and administration, medical and paramedical notes, such as hospitalization reports, radiology, EEG, endoscopy, and consultation summaries, and laboratory data.
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems, and Registries, and Biobank. Data is entered by clinicians into the hospitals EHR system, consisting of several pieces of software. Diagnoses and procedures are managed via the CORA software, drug prescription, and administration data, through the PHARMA software. Additionally, reports, radiology and consultation summaries are recorded using the AXIGATE software. These systems are integrated in the IATROS database for secondary use.
6	General representativeness	The database population are limited to patients visiting a specialised hospital.
7	Data content /source coding	Diagnoses are coded using ICD-10 and procedures are recorded using CCAM, in line with the French DRG system. Drug prescription and administration use UCD drug codes, ATC classifications, quantities, and dosages.. Additionally, medical and paramedical notes, such as hospitalization reports, radiology, EEG, endoscopy, and consultation summaries are recorded using the AXIGATE software. Laboratory data, covering both prescriptions and test results, is also included.
8	Data Harmonization	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 is assigned a unique patient number, which remains consistent throughout their care. In the rare event of duplicate records, a dedicated team is responsible for merging these records and ensuring quality control. Additionally, in France, patients have a unique national identifier (INS), which will eventually allow seamless linkage between hospital data and the SNDS (national health data system).

#	Section	Description
9	Quality control (database specific)	Each software used for data collection undergoes quality verification before allowing validation and integration into the databases. These processes are managed by the hospital's IT department. Rigorous quality control is performed at multiple stages by various stakeholders, including the IT department, the medical information department, and internal controls. Quality assurance in the source systems is managed through a series of checks. These include validation loops when studies are conducted, ensuring that data is research-ready and meets required standards. Additionally, these controls help in identifying and resolving any data inconsistencies or errors before the data is made available for research purposes.
10	Linkage	Patient-Reported Experience Measures (PREMs) and Patient-Reported Outcome Measures (PROMs) can be linked. However, this linkage is not exhaustive across all domains. While the data allows for connections between medicine usage and some health outcomes, further development is required to achieve comprehensive linkage across all patient records and conditions. In particular, using non-structured data, such as clinical notes, could be improved through Natural Language Processing (NLP) for specific conditions. The implementation of additional linkages will need to be done on a case-by-case basis. In addition, it is possible to link socioeconomic information, e.g., for indicators like the FDEP (FDep or French Deprivation Index), a measure of neighborhood deprivation.
11	Vital status	Vital status is retrieved from healthcare coverage details and covers date of death for all patients.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Fond G, Pauly V,Orleans V,Antonini F,Fabre C,Sanz M,Klay S,Jimeno MT,Leone M,Lancon C,Auquier P,Boyer L "Increased in-hospital mortality from COVID-19 in patients with schizophrenia." L'Encephale (2021): 32933762
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111141 Website: http://ap-hm.fr/

Hungary: Semmelweis University Clinical Data (SUCD)

#	Section	Description
1	Database Identification and country	SUCD (Semmelweis University Clinical Data) Budapest, Hungary
2	Data partner information section	Semmelweis University -
3	Coverage and timespan	Data collection since: 2005 Extent: Regional. The general catchment area of SU is the central region of the country, Budapest city, and Pest county, although patients can be referred from anywhere in Hungary. The total population of Budapest and Pest county is approximately 4,200,000 people. The total population of Hungary is around 9,500,000.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data, and other (specify). diagnostic data (laboratory tests, radiology, pathology)
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems, and Registries. Data is extracted directly from the source database. From there, the data entry in the system is heavily controlled and validated on the user interface before being made available for further research.
6	General representativeness	SU captures information on patients who are covered by the public health insurance system. This covers all Hungarian citizens, and therefore the database should mirror the source population well.
7	Data content /source coding	Regarding SU's source data, procedures and diagnoses are coded in SNOMED, measurements are coded in LOINC, and drugs are stored in RxNorm and ATC.
8	Data Harmonization	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 unique identifier (SSN).
9	Quality control (database specific)	The clinical database is the source database and therefore it has to be treated as a trusted database. Data entry in the systems is heavily controlled by validation on the user interface, and there are large number of rules that controls the data on the insurer's side that has to be corrected in the system by the users to be able to close the encounters. OMOP mapping is done in the framework by EHDEN recognized partners under quality check by the EHDEN society.
10	Linkage	No known linkages.
11	Vital status	Source for vital status unknown.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000184 Website: https://www.semmelweis.hu

Norway: Norwegian Linked Health Registry data (NLHR)

#	Section	Description
1	Database Identification and country	NLHR (Norwegian Linked Health Registry data) Norway
2	Data partner information section	University of Oslo Faculty of Mathematics and Natural Science – Department of Pharmacy
3	Coverage and timespan	Data collection since: 2008 Extent: Nation-wide. Norway has a universal public health care system, consisting of primary and specialist health care services covering a population of approximately 5.4 million inhabitants.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g., paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following registries are included: the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Surveillance System for Communicable Diseases (MSIS), the Norwegian Immunisation Registry (SYSVAK), the National Death Registry, and the National Registry (NR).
5	Data collection process	Registries. Many population-based health registries were established in the 1960s, with use of unique personal identifiers facilitating linkage between registries. Data in these health registries are used for health analysis, health statistics, improving the quality of healthcare, research, administration, and emergency preparedness.
6	General representativeness	The NLHR data covers the full Norwegian population.
7	Data content /source coding	NPR: ICD-10 for diagnosis, ATC and some special codes for drug use, Norwegian codes for clinical procedures (surgery (NCSP), medicine (NCMP) and diagnostic imaging, image-guided intervention, and nuclear medicine (NCRP)). KUHR: ICD-10 and ICPC-2 and ICPC-2B for diagnosis/procedure. NorPD: ATC. SYSVAK and MSIS: national classifications. MBRN: custom classifications by questionnaires (incl. check box variables in Maternity health care card)
8	Data Harmonization	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. Linkage between the registries was facilitated using project-specific person IDs generated from unique personal identification assigned at birth or immigration for all legal residents in Norway.
9	Quality control (database specific)	In-house data quality checks of rates of common conditions, drug exposures, and outcomes. We compare obtained rates with official national statistics (e.g., birth statistics, yearly rates of drug dispensing, and diagnosis by age and gender). We also review missing data and outliers and inform registry holders of any unusual patterns.
10	Linkage	The NLHR is, by definition, a linkage of datasets. Helsedata.no is one central portal to apply for 11 national health registries, including all the registries that have been mapped to the OMOP CDM.
11	Vital status	The national death registry is linked.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Mitter VR, Lupattelli A,Bjørk MH,Nordeng HME "Identification and characterization of migraine in pregnancy: A Norwegian registry-based cohort study." Cephalalgia : an international journal of headache (2024): 38663979
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000409 Website: https://www.mn.uio.no/farmasi/english/research/groups/pharma-safe/

Portugal: Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT)

#	Section	Description
1	Database Identification and country	ULSM-RT (Unidade Local de Saúde de Matosinhos Realtime Database) Distrito do Porto, Portugal
2	Data partner information section	Unidade Local de Saúde de Matosinhos Department of Research, Clinical Epidemiology and Public Health
3	Coverage and timespan	Data collection since: 2020 Extent: Regional. Complete primary and secondary public healthcare coverage for the region Matosinhos (population of 170,000 living patients as of 2023, total patients who ever lived are in excess of 200,000). Secondary healthcare coverage for the referral zone of Matosinhos (including Vila do Conde and Póvoa do Varzim) totalling an excess of 700,000 patients who ever lived.
4	Healthcare setting / type of data	Primary care – gps, and hospital inpatient care. All care settings covered are from EHR data from primary and secondary care. For procedures performed outside, claims data exists because ULSM is the payer of such procedures.
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems.
6	General representativeness	Includes patients using public sector. All socioeconomic status represented. Patients can be stratified according to high quality surrogates of socioeconomic status.
7	Data content /source coding	Medicines prescribed for outpatient pharmacy are coded natively in ATC and were translated to RxNorm at the Ingredient level. Medications prescribed from the hospital pharmacy are not coded natively. Source data terminologies used are ATC, ICD-10, ICD-9.
8	Data Harmonization	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. All patients are assigned a unique, non-repetitive number for healthcare that is the same across institutions both public and private in Portugal. There are no duplicate numbers.
9	Quality control (database specific)	Source data used represents raw EHR data as it was input by healthcare professionals for care purposes. Thus, no source of error related to data cleaning or transformations is expected to exist in source data. All quality control is done on ETL and OMOP CDM level.
10	Linkage	No known linkages.
11	Vital status	Source for vital status unknown.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111171 Website: https://www.ulsm.min-saude.pt

Spain: Platafoma de Recerca en Informació Sanitària de les Illes Balears (PRISIB)

#	Section	Description
1	Database Identification and country	PRISIB (Platafoma de Recerca en Informació Sanitària de les Illes Balears) Balearic Islands, Spain
2	Data partner information section	IdISBa Health Data Research Platform of the Balearic Islands
3	Coverage and timespan	Data collection since: 2018 Extent: Regional. The geographic area of catchment for the PRISIB database includes all of the Balearic Islands (Mallorca, Menorca, Ibiza, Formentera, etc.). This is estimated to encompass the whole population of the archipelago. The approximate population is around 1245000 inhabitants.

#	Section	Description
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g., paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The PRISIB database includes hospital data comprising visits, measurements, diagnoses, procedures, laboratory results as well as medication prescriptions.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. Electronic system for outpatient's drug prescriptions.
6	General representativeness	There are no socioeconomic specific characteristics of the database and should therefore mirror the source population well.
7	Data content /source coding	The following terminology systems are used: ATC, ICD9, ICD10, ICPC-2, National drug agency catalogue of products and the Spanish radiology association catalogue.
8	Data Harmonization	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. No.
9	Quality control (database specific)	When a dataset is generated from this source, we use a script that validates that what we extracted conforms to the criteria stated in the data model for each particular project and generates a descriptive and data quality report.
10	Linkage	No known linkages.
11	Vital status	It is assumed that the database only captures in-hospital deaths or deaths occurring in the geographic area.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: Website: https://www.idisba.es/en/Support-Services/Scientific-Technical-Platforms/Research-in-Health-Information

Spain: Valencia Health System Integrated Dataset (VID)

#	Section	Description
1	Database Identification and country	VID (Valencia Health System Integrated Dataset) Comunitat Valenciana, Spain
2	Data partner information section	FISABIO Health Services Research & Pharmacoepidemiology Unit
3	Coverage and timespan	Data collection since: 2008 Extent: Regional. The VID covers the general population of the Valencia region, comprising 10.7% of the Spanish population. The total population is estimated to be around 5,300,000.
4	Healthcare setting / type of data	Primary care – gps, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify). Both primary and secondary care settings are covered, where visits, diagnoses, medications, measurements, and procedures are recorded. The population information system collects demographics data. The Electronic prescription and dispensing system capture medication, while the Minimum Basic Data set records all admissions, linked diagnoses, and procedures. Measurements are captured additionally from the vaccine and microbial disease registries.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. Data extraction is performed by clinical IT personnel. Data is released by the health authorities on a project basis and can only be used for such purposes.

#	Section	Description
6	General representativeness	The population captured by the VID should represent the Valencia region well, as the VID contains data of the general population covered by the universal public health care system. About 97% of the population in this region is covered by public care.
7	Data content /source coding	Prescribed and dispensed medications are coded with the ATC system. The indications of each prescription, as well as procedures are coded using ICD9CM and ICD10ES.
8	Data Harmonization	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. No.
9	Quality control (database specific)	The data is reviewed carefully with the IT personnel who perform the extraction of data and then by a senior researcher with expertise in RWD management in the HSRP unit. Several quality check scripts are run against the received data. Finally, a senior researcher with RWD and clinical expertise assesses the completeness, consistency, and quality of the data extraction. If any inconsistency or error is detected, the dataset is requested and extracted again.
10	Linkage	VID also contains hospital discharge records, emergency care discharge records, birth registry, congenital anomaly registry, perinatal mortality registry, cancer registry, pharmacy prescription and dispensing records, vaccine records, and microbiology records. Most databases are updated daily, but certain registries, such as the metabolic disease and perinatal mortality registries, are updated monthly and yearly respectively.
11	Vital status	Source for vital status unknown.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	García-Sempere A, Orrico-Sánchez A,Muñoz-Quiles C,Hurtado I,Peiró S,Sanfélix-Gimeno G,Diez-Domingo J "Data Resource Profile: The Valencia Health System Integrated Database (VID)." International journal of epidemiology (2020): 31977043
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111174 Website: http://www.sp.san.gva.es/

ANNEX II. Fitness for use assessment

This is part of a larger study which includes three studies (P4-C1-022, P4-C2-015, and P4-C2-016). All three studies share the same objectives but differ in the data partners involved due to the need to assess how suicide, suicide-related events, and depression are captured across the majority of the DARWIN EU® network. The three protocols are used due to the number of data sources included. As a consequence, the three studies overall include data partners who constitute the majority of the DARWIN EU® network, with a few exceptions: neonatal data sources (The United Kingdom: National Neonatal Research Database (NNRD)) and disease specific registries (The Netherlands: Netherlands Cancer Registry (NCR), Norway: Cancer Registry Norway (CRN), Spain: HARMONY - Acute Lymphoblastic Leukemia (ALL), HARMONY - Acute Myeloid Leukemia (AML), HARMONY - Chronic Myeloid Leukemia (CML), and HARMONY - Multiple Myeloma (MM)), as they can be considered less relevant and therefore out of scope in regards to the study objectives. One data source is currently unavailable to execute DARWIN EU® studies (France: Système National des Données de Santé (SNDS)) and is therefore not included in any of the studies. Other data sources, including CDWBordeaux, IQVIA DA Germany, IQVIA LPD Belgium, EMDB-ULSEDV, UKBB, POLIMI, and FinOMOP-ACI Varha declined involvement due to unsuitable data or were considered to have insufficient counts expected.

Table S1. Overview of data sources included in all studies: P4-C1-022, P4-C2-015, and P4-C2-016.

Country	Data partner	Database type	Study
Croatia	NAJS	Registry	P4-C1-022
Denmark	DK-DHR	Registry	P4-C1-022
Estonia	EBB	Registry	P4-C1-022
Germany	InGef RDB	Claims	P4-C1-022
The Netherlands	IPCI	Outpatient General Practitioner Care	P4-C1-022
The United Kingdom	CPRD GOLD	Outpatient General Practitioner Care	P4-C1-022
Greece	PGH	Inpatient (Hospital) Care	P4-C2-015
Portugal	EMDB-ULSGE	Inpatient (Hospital) Care	P4-C2-015
Portugal	EMDB-ULSRA	Inpatient (Hospital) Care	P4-C2-015
Spain	BIFAP	Outpatient general practitioner And Hospital Care	P4-C2-015
Spain	H12O	Inpatient And Outpatient Hospital Care	P4-C2-015
Spain	IMASIS	Inpatient (Hospital) Care	P4-C2-015
Spain	SIDIAP	Outpatient General Practitioner Care	P4-C2-015
Sweden	HI-SPEED	Registry, Outpatient General Practitioner Care, Inpatient (Hospital) Care	P4-C2-015
Finland	FinOMOP-THL	Registry	P4-C2-016
Finland	FinOMOP-HUS	Inpatient (Hospital) Care	P4-C2-016
Finland	FinOMOP-TaUH Pirha	Inpatient (Hospital) Care	P4-C2-016
France	APHM	Inpatient And Outpatient Hospital Care	P4-C2-016
Hungary	SUCD	Inpatient (Hospital) Care	P4-C2-016
Norway	NLHR	Registry	P4-C2-016
Portugal	ULSM-RT	Inpatient (Hospital) Care	P4-C2-016
Spain	PRISIB	Outpatient General Practitioner Care, Inpatient (Hospital) Care	P4-C2-016

Country	Data partner	Database type	Study
Spain	VID	Outpatient General Practitioner Care	P4-C2-016

Acronyms: NAJS = Croatian National Public Health Information System, DK-DHR = Danish Data Health Registries, EBB = Estonian Biobank, FinOMOP THL = Finnish Care Register for Health Care, FinOMOP HUS = Hospital District of Helsinki and Uusimaa, FinOMOP-TaUH Pirha = Tampere University Hospital patient cohort, APHM = Assistance Publique Hôpitaux de Marseille, InGef = Institute for Applied Health Research Berlin GmbH Research Database, PGH = Papageorgiou General Hospital, SUCD = Semmelweis University Clinical Data, IPCI = Integrated Primary Care Information, NLHR = Norwegian Linked Health Registry data, EMDB-ULSGE = Egas Moniz Health Alliance database - Gaia E Espinho, EMDB-ULSRA = Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro), ULSM-RT = Unidade Local de Saúde de Matosinhos Realtime Database, BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, H12O = Hospital Universitario 12 de Octubre, IMASIS = Institut Municipal Assistència Sanitària Information System, PRISB = Platafoma de Recerca en Informació Sanitària de les Illes Balears, SIDIAP = The Information System for the Development of Research in Primary Care, VID = Valencia Health System Integrated Dataset, HI-SIPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD = Clinical Practice Research Datalink GOLD

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® Remote Research Environment (RRE). These output files do not contain any data that allow identification of subjects included in the study. The RRE 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

The *CohortDiagnostics* (https://github.com/OHDSI/CohortDiagnostics) R package 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, and *CohortCharacteristics* to characterise the cohort. 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 European Medicines Agency (EMA) by the DARWIN EU® Coordination Centre (CC) upon completion of the study.

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

ANNEX IV. List of stand-alone documents

Table S2. Preliminary list of condition and observation definitions.

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Suicide	Suicide	440925	N/A	SNOMED
Composite suicide-related events (suicide, suicide attempt, suicide ideation, self-harm)	At increased risk for suicide Feeling suicidal Harmful thoughts Injury due to suicide attempt Intentional overdose Intentional poisoning by drug Intentionally harming self Late effect of self inflicted injury Self destructive behaviour Self inflicted injury Self-administered poisoning Suicidal poisoning Suicide Suicide attempt Suicide plan Suicide risk Threatening suicide	4021336 4021339 4037303 4257906 607149 4196024 4303690 435446 608248 439235 4181216 444362 440925 4219484 600767 37399733 4216115	4190444 42573949 42573140 42536693 4206010 42596336	SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED
Suicide attempt	Injury due to suicide attempt Intentional overdose Intentional poisoning by drug Self-administered poisoning Suicidal poisoning Suicide attempt	4257906 607149 4196024 4181216 444362 4219484	4206010	SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED
Suicide ideation	At increased risk for suicide Feeling suicidal Harmful thoughts Suicide plan Suicide risk Threatening suicide	4021336 4021339 4037303 600767 37399733 4216115	4190444	SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED
Self-harm	Intentional overdose Intentional poisoning by drug Intentionally harming self Late effect of self inflicted injury Self destructive behaviour Self inflicted injury Self-administered poisoning Suicidal poisoning	607149 4196024 4303690 435446 608248 439235 4181216 444362	42573949 42573140 42536693 4206010 42596336 440925	SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED
Depression	Depressive disorder Depressive mood	440383 40546087	N/A N/A	SNOMED SNOMED
Schizophrenia	Schizophrenia	435783	N/A	SNOMED
Anxiety disorder	Anxiety disorder	442077	N/A	SNOMED
Substance abuse (excluding tobacco and alcohol)	Substance use disorder	1448778	435243 1448827 4218106 378726 433753 37110444 35610532 4209423	SNOMED

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
			1448842 437264	
Alcohol abuse	Alcohol use disorder Dementia associated with alcoholism Harmful use of alcohol	1448827 378726 35610532	439005	SNOMED SNOMED SNOMED SNOMED
Personality disorder	Personality disorder	441838	N/A	SNOMED
Bipolar disorder	Bipolar disorder	436665	N/A	SNOMED
Eating disorder	Eating disorder	439002	4143677 46285098	SNOMED
Resolution/outcome of suicide-related events and depression	At low risk for suicide Depression resolved Depressive disorder in remission No apparent risk of suicide No suicidal thoughts	4190444 44788804 44782943 44809512 4046664	N/A N/A N/A N/A N/A	SNOMED SNOMED SNOMED SNOMED SNOMED
Smoking	Cigarette smoker Current smoker Documentation of smoking cessation medication action plan Harmful pattern of use of nicotine Negotiated date for cessation of smoking Nicotine dependence Nicotine use disorder Referral to smoking cessation service offered Referral to smoking cessation advisor Smoking cessation assistance Smoking cessation program start date Stop smoking monitoring check done Tobacco dependence syndrome Tobacco or its derivatives user	903657 40766945 762882 37110444 4193014 4209423 1448842 44804580 4217594 4293153 4269046 4090849 437264 903654	2108437	OMOP Extension LOINC SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED OMOP Extenstion
Hospital admission	Admission by health worker Admission for care Admission to department Admission to high secure unit Admission to hospice Admission to inpatient rehabilitation facility Admission to medium secure unit Admission to regional secure unit Admission under the Mental Health Act Emergency Room and Inpatient Visit Hospital admission Inpatient Visit Involuntary admission Listed for admission to hospital Referral to physician Voluntary admission	4137737 4085709 4137915 44787815 4123927 763992 44812023 44811987 44790848 262 8715 9201 4046990 4084992 4084682 4047815	2110331 2110332 2110333 2110334 2110335 2110336	SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED Visit SNOMED Visit SNOMED SNOMED SNOMED SNOMED
Psychiatric hospital admission	Admission to psychiatric day hospital Admission to psychiatric intensive care unit Admission to psychiatry department Admission to psychogeriatric day hospital Emergency hospital admission to forensic psychiatry service Emergency hospital admission to perinatal	4138932 44812116 4138051 4123931 35609102 35609156 35609155	N/A	SNOMED SNOMED SNOMED SNOMED SNOMED SNOMED

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
	psychiatry service Emergency hospital admission to psychotherapy service Inpatient psychiatric facility Listed for psychiatric admission Psychiatric emergency hospital admission Psychiatric facility-partial hospitalization Psychiatric hospital Psychogeriatric emergency hospital admission	8971 4084994 4084664 8913 38004284 4079625		SNOMED SNOMED CMS Place of Service SNOMED SNOMED CMS Place of Service SNOMED SNOMED
Hospital referrals	Referral to doctor Referral to hospital Subject requests referral	4127751 4147710 40480414	4205379 4167404 4081621 44809310 4141716 36675630 4084358 44790931 44805697 2721126	SNOMED SNOMED SNOMED
Obesity	Obese	4215968	N/A	SNOMED
Pain	Pain	4329041	N/A	SNOMED
Chronic pain	Chronic pain	436096	N/A	SNOMED
Cancer (including haematological and excluding melanoma)	Carcinoma in situ Malignant neoplastic disease Neoplasm of hematopoietic cell type	433435 443392 4189640	N/A N/A N/A	SNOMED SNOMED SNOMED
PHQ-9	Mood interview total severity score [Reported PHQ-9 CMS] Patient Health Questionnaire 9 item (PHQ-9) total score [reported] PHQ-9 – Patient health questionnaire 9 Positive screening for depression on PHQ-9 (Patient Health Questionnaire 9)	40757785 3042932 44804610 430211839	N/A N/A N/A N/A	LOINC LOINC SNOMED SNOMED
Beck Depression Inventory (I and II)	Beck Depression Inventory Beck Depression Inventory Fast Screen [BDI] Beck Depression Inventory Fast Screen total score [BDI] Beck Depression Inventory II Beck Depression Inventory II [BDI] Beck Depression Inventory II total score [BDI]	4167608 36305762 36303939 1447419 36304859 36305191	N/A N/A N/A N/A N/A N/A	SNOMED LOINC LOINC SNOMED LOINC LOINC
Hamilton depression scale	Hamilton rating scale for depression	4159709	N/A	SNOMED
Columbia Suicide Severity Scale	Columbia Suicide Severity Rating Scale Columbia - suicide severity rating scale screener - recent [C-SSRS] Columbia - suicide severity rating scale - very young child or cognitively impaired - lifetime recent [C-SSRS] Columbia - suicide severity rating scale - lifetime recent [C-SSRS]	3655548 1002134 1002203 1002447	N/A N/A N/A N/A	SNOMED LOINC LOINC LOINC

Table S3. Preliminary list of medicines definitions.

Substance Name	ATC code	Concept name	Ingredient Concept ID	Include descendants
Antidepressants	N06AA	5-hydroxytryptophan	1363516	Yes*
	N06AB	Agomelatine	36878783	
	N06AF	Alaprocate	36856226	
	N06AG	Amineptin	19032424	
	N06AX	Amitriptyline	710062	
		Amoxapine	713109	
		Bifemelane	19058055	
		Brexanolone	1510996	
		Bupropion	750982	
		Butriptyline	19039883	
		Citalopram	797617	
		Clomipramine	798834	
		Desipramine	716968	
		Desvenlafaxine	717607	
		Dibenzepin	19023846	
		Dimetacrine	40798767	
		Dothiepin	19037989	
		Doxepin	738156	
		Duloxetine	715259	
		Esketamine	715939	
		Etoperidone	1366610	
		Fluoxetine	36858847	
		Fluvoxamine	755695	
		Gepirone	751412	
		Hypericum extract	36858670	
		Hypericum perforatum leaf extract	42899139	
		Imipramine	42899140	
		Iprindole	778268	
		Iproclozide	19122204	
		Iproniazid	36849601	
		Isocarboxazid	19122259	
		Levomilnacipran	781705	
		Lofepramine	43560354	
		Maprotiline	19091830	
		Medifoxamine	794147	
		Melitracen	19072021	
		Mianserin	19110751	
		Milnacipran	19007737	
		Minaprine	19080226	
		Mirtazapine	19072123	
		Moclobemide	725131	
		Nefazodone	19010652	
		Nialamide	714684	
		Nomifensine	19127550	
		Nortriptyline	19128066	
		Opipramol	721724	
		Oxaflozane	19129635	
		Paroxetine	36862328	
		Phenelzine	722031	
		Pivagabine	733896	
		Protriptyline	36856232	
		Quinupramine	754270	
		Reboxetine	19098514	
		Sertraline	19084693	
		St. John's wort extract	739138	
		Tianeptine	1398039	
		Toloxatone	19041910	
		Tranylcypromine	19100775	
		Trazodone	703470	

Substance Name	ATC code	Concept name	Ingredient Concept ID	Include descendants
		Trimipramine	703547	
		Tryptophan	705755	
		Venlafaxine	19006186	
		Vilazodone	743670	
		Viloxazine	40234834	
		Vortioxetine	19008261	
		Zimeldine	44507700	
			40799195	
Antipsychotics	N05AA	Acepromazine	19018226	Yes*
	N05AB	Acetophenazine	19029555	
	N05AC	Amisulpride	19057607	
	N05AD	Aripiprazole	757688	
	N05AE	Asenapine	40164052	
	N05AF	Benperidol	19016440	
	N05AG	Brexpiprazole	46275300	
	N05AH	Bromperidol	19039227	
	N05AL	Butaperazine	40798666	
	N05AN	Cariprazine	35603277	
	N05AX	Chlorproethazine	19122262	
		Chlorpromazine	794852	
		Chlorprothixene	19095002	
		Clopenthixol	36848877	
		Clothiapine	19100363	
		Clozapine	800878	
		Cyamemazine	19051234	
		Dixyrazine	40798772	
		Droperidol	739323	
		Fluanisone	40798823	
		Flupenthixol	19055982	
		Fluphenazine	756018	
		Fluspirilene	19056465	
		Haloperidol	766529	
		Iloperidone	19017241	
		Levosulpiride	43009023	
		Lithium	19124477	
		Loxapine	792263	
		Lumateperone	37498659	
		Lurasidone	40230761	
		Mesoridazine	703083	
		Methotrimeprazine	19005147	
		Metylperon	19072088	
		Molindone	709699	
		Moperone	40798964	
		Mosapramine	36848724	
		Mosapramine hydrochloride	35198101	
		Olanzapine	785788	
		Oxypertine	19025922	
		Paliperidone	703244	
		Penfluridol	19028044	
		Perazine	19131663	
		Periciazine	19053565	
		Perphenazine	733008	
		Pimavanserin	42628962	
		Pimozide	745790	
		Pipamperone	19093225	
		Pipothiazine	19133992	
		Prochlorperazine	752061	
		Promazine	19052903	
		Prothipendyl	19115044	
		Quetiapine	766814	

Substance Name	ATC code	Concept name	Ingredient Concept ID	Include descendants
		Remoxipride	19035226	
		Risperidone	735979	
		Sertindole	19050633	
		Sulpiride	19136626	
		Sultopride	19100431	
		Thiopropazate	19041817	
		Thioproperazine	19000305	
		Thioridazine	700299	
		Thiothixene	700465	
		Tiapride	19008012	
		Trifluoperazine	704984	
		Trifluperidol	19005101	
		Triflupromazine	19005104	
		Veralipride	19043327	
		Ziprasidone	712615	
		Zotepine	19102109	
		Zuclopenthixol	19010886	
Benzodiazepines	N03AE N05BA N05CD	Adinazolam	36858914	Yes*
		Alprazolam	781039	
		Bentazepam	36852894	
		Bromazepam	19030353	
		Brotizolam	19039262	
		Camazepam	40798714	
		Chlordiazepoxide	990678	
		Cinolazepam	36852970	
		Clobazam	19050832	
		Clonazepam	798874	
		Clorazepate	790253	
		Clotiazepam	19068821	
		Cloxazolam	19051096	
		Diazepam	723013	
		Doxefazepam	36849918	
		Estazolam	748010	
		Ethyl loflazepate	19095467	
		Etizolam	43009000	
		Fludiazepam	35198139	
		Flunitrazepam	19055224	
		Flurazepam	756349	
		Halazepam	801396	
		Ketazolam	19003946	
		Lorazepam	791967	
		Lormetazepam	19007977	
		Medazepam	19125106	
		Mexazolam	43009058	
		Midazolam	708298	
		Nimetazepam	35197946	
		Nitrazepam	19020021	
		Nordazepam	19080959	
		Oxazepam	724816	
		Pinazepam	2100521	
		Prazepam	19050461	
		Quazepam	731188	
		Remimazolam	1146633	
		Temazepam	836715	
		Tofisopam	19100773	
		Triazolam	704599	
		Triazulenone	19042550	

Substance Name	ATC code	Concept name	Ingredient Concept ID	Include descendants
Stimulants	N06BA N06BC N06BX	Acetylcarnitine Adrafinil Amphetamine Aniracetam Armodafinil Atomoxetine Caffeine Citicoline Deanol Dexmethylphenidate Dextroamphetamine Fencamfamin Fenethylline Fenozolone Glycofurol Idebenone Lisdexamfetamine Meclofenoxate Methamphetamine Methylphenidate Modafinil Nizofenone fumarate Oxiracetam Pemoline Pipradrol Hydrochloride Piracetam Prolintane Propentofylline Pyrisuccideanol Pyrithioxin Solriamfetol Vinpocetine	19037596 19112518 714785 19032667 19090984 742185 1134439 40223464 1311826 731533 719311 40798833 19061288 19061392 43532437 19067036 709567 19052682 704053 705944 710650 35198179 19082891 727835 44012716 19046654 19029018 19096031 19097999 19060358 1510970 19095462	Yes*
Hypnotics	NO5CA NO5CB NO5CC NO5CE NO5CF NO5CH NO5CJ NO5CM NO5CX	Allobarbital Allylisopropylacetylurea Ammonium bicarbonate Amobarbital Anise oil Aprobarbital Barbital Bromisoval Butobarbital Butylvinal Calcium bromide Carbromal Chloral betaine Chlormethiazole Chloroform Cyclobarbital Daridorexant Dexmedetomidine Dichloralphenazone Eszopiclone Ethanol Ethchlorvynol Glutethimide Glycyrrhiza glabra extract Heptabarb Hexobarbital Lemborexant Melatonin	19113034 35197991 36878870 712757 19060831 19030299 19015346 40798676 19039767 19021512 19018639 19033243 19135825 19092283 43012199 19072082 1758995 19061088 1189490 757352 955372 749727 19061124 43013634 40798884 19068964 37498491 1301152	Yes*

Substance Name	ATC code	Concept name	Ingredient Concept ID	Include descendants
		Methaqualone	19004160	
		Methohexital	19005015	
		Methylpentynol	40798942	
		Methyprylon	19006742	
		Niaprazine	19082402	
		Paraldehyde	19027181	
		Pentobarbital	730729	
		Potassium bromide	19095079	
		Propiomazine	19053610	
		Proxibarbal	19134133	
		Pyrithyldione	40799066	
		Ramelteon	781182	
		Scopolamine	965748	
		Secobarbital	766067	
		Sodium bromide	42900071	
		Suvorexant	45775760	
		Talbutal	19059149	
		Tasimelteon	44814600	
		Thiopental	700253	
		Trichloroacetaldehyde	742594	
		Triclofos	19042692	
		Valerian root extract	1397059	
		Valeriana officinalis whole extract	44818506	
		Vinbarbital	19135827	
		Zaleplon	720727	
		Zolpidem	744740	
		Zopiclone	19044883	

*All topical concepts were excluded

Table S4. Preliminary list of procedures definitions.

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Psychotherapy	Psychotherapy	4327941	N/A	SNOMED
Electrotherapy	Electroconvulsive therapy Transcranial direct current stimulation Transcranial magnetic stimulation	4030840 37158172 4130731	N/A N/A N/A	SNOMED SNOMED SNOMED

ANNEX V. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Characterisation of data capture related to suicidality and depression across the DARWIN EU® network

EU PAS Register® number: EUPAS1000000825
Study reference number (if applicable): P4-C1-022/P4-C2-015/P4-C2-016

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]	☒	☐	☐	5
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.	☒	☐	☐	5

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	7
2.1.1 Why the study is conducted? (e.g., to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	6
2.1.2 The objective(s) of the study?	☒	☐	☐	7
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	8.1
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

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

[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**
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	8.4
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	8.8
3.4 Does the protocol specify measure(s) of association? (e.g., risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☒	
3.5 Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g., adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	☒	☐	☐	8.3
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	☒	☐	☐	8.3
4.2.2 Age and sex	☒	☐	☐	8.3
4.2.3 Country of origin	☒	☐	☐	8.3
4.2.4 Disease/indication	☒	☐	☐	8.3
4.2.5 Duration of follow-up	☒	☐	☐	8.3
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)	☐	☐	☒	
5.2 Does the protocol address the validity of the exposure measurement? (e.g., precision, accuracy, use of validation sub-study)	☐	☐	☒	
5.3 Is exposure categorised according to time windows?	☐	☐	☒	
5.4 Is intensity of exposure addressed? (e.g., dose, duration)	☐	☐	☒	

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

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	8.6.2
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.6.2
6.3 Does the protocol address the validity of outcome measurement? (e.g., precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☒	☐	
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g., HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g., confounding by indication)	☒	☐	☐	8.6, 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)	☒	☐	☐	9

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)	☒	☐	☐	8.33, 8.6.4, 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)	☒	☐	☐	8.6.1
9.1.2 Outcomes? (e.g., clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	8.6.2
9.1.3 Covariates and other characteristics?	☒	☐	☐	8.6.4
9.2 Does the protocol describe the information available from the data source(s) on:				8
9.2.1 Exposure? (e.g., date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☒	☐	☐	8.6.1
9.2.2 Outcomes? (e.g., date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	8.6.2
9.2.3 Covariates and other characteristics? (e.g., age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	8.6.4
9.3 Is a coding system described for:				8
9.3.1 Exposure? (e.g., WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	8.6.1
9.3.2 Outcomes? (e.g., International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	8.6.2
9.3.3 Covariates and other characteristics?	☒	☐	☐	8.6.4
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?	☒	☐	☐	8.8
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.7 Does the plan describe methods for handling missing data?	☒	☐	☐	8.8
10.8 Are relevant sensitivity analyses described?	☐	☐	☒	

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?	☒	☐	☐	Annex III

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?	☒	☐	☐	Annex II

Comments:

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

Comments:

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

Comments:

Name of the main author of the protocol: Nicholas Hunt

Date: 03/12/2025

Signature: 

ANNEX VI. Glossary

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

Aggregated Data

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

Benefit-Risk Assessment

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

Common Data Model (CDM)

A 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.