



Study Protocol

P5-C1-006

DARWIN EU® - Sex differences in treatment patterns for venous thromboembolic events in European countries

22/09/2026

Version 6.0

Authors: Amy Lam, Nuria Mercade Besora, Annika Jödicke

Public

CONTENTS

LIST OF ABBREVIATIONS	5
1. TITLE	7
2. DESCRIPTION OF THE STUDY TEAM	7
3. ABSTRACT	8
4. AMENDMENTS AND UPDATES	11
5. MILESTONES	11
6. RATIONALE AND BACKGROUND	11
7. RESEARCH QUESTION AND OBJECTIVES	12
8. RESEARCH METHODS	12
8.1. Study design	12
Figure 1. Graphical depiction of the study design of Objective 1.	13
Figure 2. Graphical depiction of the study design of Objective 2.	14
Figure 3. Graphical depiction of the study design of Objective 3.	15
8.2. Follow-up	15
8.3. Study population with inclusion and exclusion criteria	16
8.4. Study setting and data sources	16
Table 1. Data sources	16
8.5. Study period	17
8.6. Variables	17
8.6.1. Exposure	17
Table 2. Pre-specified antithrombotics for ingredient-level analyses	18
Table 3. Pre-specified antithrombotics for drug class-level analyses.	19
8.6.2. Covariates, including confounders, effect modifiers, and other variables	19
8.7. Study size	20
8.8. Analysis	20
8.8.1. Federated network analyses	20
8.8.2. Data privacy protection	20
8.8.3. Statistical model specification and assumptions of the analytical approach considered	21
8.8.4. Output	22
8.9. Evidence synthesis	22
9. STRENGTHS AND LIMITATIONS	22
10. REFERENCES	24
11. ANNEXES	25
ANNEX I. Description of data sources	25
Denmark: Danish Data Health Registries (DK-DHR)	25
Finland: Finnish Care Register for Health Care (FinOMOP-THL)	26
Spain: The Information System for Research in Primary Care (SIDIAP)	28
The Netherlands: Integrated Primary Care Information (IPCI)	29
The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)	30
ANNEX II. Fitness for use assessment	33
ANNEX III. Operational and reporting considerations	35
ANNEX IV. List of concepts	37
Table S1. Preliminary list of condition concept IDs for defining VTE.	37

ANNEX V. Feasibility counts.....	41
Table S2. Descendant Patient Counts of antithrombotics (ATC B01) in participating data sources. ...	41
ANNEX VI. Mock tables and figures.....	43
Table S3. Mock Table: Characteristics of baseline comorbidities and initial treatment of individuals with VTE, in men and women.....	43
Table S4. Mock Table: Characteristics of baseline comorbidities and initial treatment of individuals with VTE, in men and women, stratified by age group.	45
Table S5. Mock Table: Drug utilisation of initial treatment for VTE, in men and women.....	52
Figure S1. Mock Figure: Sankey plot for treatment switching patterns at drug-class level in men, per data source.	54
Figure S2. Mock Figure: Sankey plot for treatment switching patterns at drug-class level in women, per data source.....	55
Figure S3. Mock Figure: Sankey plot for treatment switching patterns at ingredient level in men, per data source.	56
Figure S4. Mock Figure: Sankey plot for treatment switching patterns at ingredient level in women, per data source.....	57
ANNEX VII. ENCePP checklist for study protocols	58
ANNEX VIII. Glossary.....	64

Study title	DARWIN EU® - Sex differences in treatment patterns for venous thromboembolic events in European countries
Protocol version	V6.0
Date	22/09/2026
EUPAS number	EUPAS1000001079
Active Substance	Antithrombotics (ATC code B01)
Medicinal Product	Not applicable
Research question and objectives	<p><u>Research question:</u></p> <p>To what extent do treatment patterns for thromboembolic disease differ between men and women in clinical practice in European countries?</p> <p><u>Study objectives:</u></p> <ol style="list-style-type: none"> 1. To describe antithrombotic treatment initiation for venous thromboembolic events in men and women 2. To describe drug utilisation of antithrombotic treatments, including initial treatment duration and initial dose, in men and women 3. To describe antithrombotic treatment switching patterns for venous thromboembolic events in men and women <p>All analyses will be conducted overall and stratified by age groups, pre-existing conditions, and use of hormonal contraception/hormone replacement therapy (HRT) [in women only].</p>
Countries of study	Denmark, Finland, Spain, the Netherlands, the United Kingdom
Authors	<p>Amy Lam a.lam@darwin-eu.org</p> <p>Núria Mercadé Besora n.mercadebesora@darwin-eu.org</p> <p>Annika Jödicke a.jodicke@darwin-eu.org</p>

LIST OF ABBREVIATIONS

Acronyms/terms	Description
ATC	Anatomical Therapeutic Chemical [classification]
BMI	Body Mass Index
CC	Coordination Centre
CDM	Common Data Model
CPRD GOLD	Clinical Practice Research Datalink GOLD
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
DKMA	Danish Medicines Agency
DRE	Digital Research Environment
DTZ	Data Transfer Zone
DVT	Deep Vein Thrombosis
EHR	Electronic Health Records
EMA	European Medicines Agency
EMC	Erasmus University Medical Center
ENCEPP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
FinOMOP-THL	Finnish Care Register for Health Care
GDPR	General Data Protection Regulation
GP	General Practitioner
HRT	Hormone replacement therapy
ICD	International Classification of Diseases
IDIAPJGol	Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina
IP	Inpatient
IPCI	Integrated Primary Care Information
IQR	Interquartile range
IRB	Institutional Review Board
LOINC	Logical Observation Identifiers Names and Codes
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
Oxford NDORMS	University of Oxford - Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences
PE	Pulmonary Embolism
RxNorm	Medical prescription normalised
SIDIAP	The Information System for Research in Primary Care

Acronyms/terms	Description
SNOMED	Systematized Nomenclature of Medicine
VTE	Venous thromboembolism
WHO	World Health Organization

1. TITLE

DARWIN EU® - Sex differences in treatment patterns for venous thromboembolic events in European countries

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Ivan Lam	University of Oxford
	Annika Jodicke	University of Oxford
Data Scientist	Nuria Mercade Besora	University of Oxford
	Edward Burn	University of Oxford
Clinical Domain Expert	Albert Prats Uribe	University of Oxford
	Daniel Prieto-Alhambra	University of Oxford, Erasmus MC
Epidemiologist	Xintong Li	University of Oxford
	Amy Lam	University of Oxford
Study Manager	Natasha Yefimenko Nosova	Erasmus MC
Data source	Names	Data Partner Organisation*
Danish Data Health Registries (DK-DHR)	Elvira Bräuner	Danish Medicines Agency (DKMA)
	Susanne Bruun	
Finnish Care Register for Health Care (FinOMOP-THL)	Toni Lehtonen	FinOMOP
The Information System for Research in Primary Care (SIDIAP)	Elena Roel Herranz	Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
	Augustina Giuliadori Picco	
	Laura Granés	
	Talita Duarte Salles	
Integrated Primary Care Information (IPCI)	Katia Verhamme	Erasmus University Medical Center (EMC)
	Mees Mosseveld	
	Marcel de Wilde	
	Tineke de Ben	
Clinical Practice Research Datalink GOLD (CPRD GOLD)	Antonella Delmestri	University of Oxford - Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences (Oxford NDORMS)

*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® - Sex differences in treatment patterns for venous thromboembolic events in European countries

Rationale and background

Antithrombotic agents are used in the management of venous thromboembolism (VTE). Current guidelines recommend direct oral anticoagulants, heparins, and vitamin K antagonists for the prophylaxis of primary and secondary VTE, whilst fibrinolytic agents are used for acute dissolving of clots in Deep Vein Thrombosis (DVT) and Pulmonary Embolism (PE). Optimal treatment may differ between men and women due to biological and clinical factors, including bleeding risk, pharmacokinetics, and hormonal influences, such as oral contraceptives and hormone replacement therapy (HRT). Important evidence gaps remain regarding sex-specific antithrombotic treatment strategies, particularly in the management of women receiving hormonal contraception or HRT during anticoagulant therapy.

As part of the broader European Medicines Agency (EMA) Women's Health Initiative, this study aims to evaluate potential sex differences in the treatment of VTE. By examining sex-specific patterns in treatment initiation and subsequent treatment pathways, the study seeks to identify potential gaps in care that may inform clinical practice, regulatory decision-making, and public health strategies aimed at promoting equitable, evidence-based treatment.

Research question and objectives

Research question

To what extent do treatment patterns for thromboembolic disease differ between men and women in clinical practice in European countries?

Objectives

Specific objectives:

1. To describe antithrombotic treatment initiation for venous thromboembolic events in men and women
2. To describe drug utilisation of antithrombotic treatments, including initial treatment duration and initial dose, in men and women
3. To describe antithrombotic treatment switching patterns for venous thromboembolic events in men and women

All analyses will be conducted overall and stratified by age groups, pre-existing conditions, and use of hormonal contraception/HRT [in women only].

Methods

Study design

A (new user) cohort study will be conducted using routinely collected health data from 5 data sources from 5 countries across Europe.

Population

Objective 1 (Characterisation of antithrombotic treatment initiation)

Inclusion criterion: First recorded diagnosis of VTE during the recruitment period

Exclusion criteria

- Age <18 years at first recorded diagnosis of VTE (index date)
- Less than 365 days of available history before the index date
- Record of VTE before study start (01/01/2015)
- Record of prescription/dispensation of antithrombotics before the index date

Objectives 2 and 3 (Drug utilisation and switching patterns)

Inclusion criteria

- First recorded diagnosis of VTE during the recruitment period
- Initiation of antithrombotic drug treatment within 90 days after diagnosis (index date)

Exclusion criteria

- Age <18 years on the date of antithrombotic treatment initiation
- Less than 365 days of available history before date of antithrombotic treatment initiation
- Record of VTE before study start 01/01/2015
- Record of antithrombotic drug prescription/dispensation before first recorded diagnosis of VTE

Variables

Exposure:

Exposure to antithrombotic treatment will be identified as the first prescription or dispensing recorded after VTE. Antithrombotic drug treatment will be defined as ingredients classified under Anatomical Therapeutic Chemical [classification] (ATC) B01 “Antithrombotic treatments”:

- B01AA – Vitamin K antagonist
- B01AB – Heparin
- B01AC – Platelet aggregation inhibitors
- B01AD – Enzymes
- B01AE – Direct thrombin inhibitors
- B01AF – Direct factor Xa inhibitors
- B01AX – Other antithrombotic agents

Drug class level analyses will include all 7 ATC groups, whereas ingredient-level analyses will exclude platelet aggregation inhibitors. Based on feasibility counts for the participating data sources, a total of 17 antithrombotic drugs have been pre-selected for ingredient-level analyses.

Covariates:

Age, age group, sex, baseline comorbidities (chronic kidney disease, active cancer, extreme under/overweight), use of hormonal contraception/HRT [in women only].

Data sources

1. Denmark: Danish Data Health Registries (DK-DHR)
2. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
3. Spain: The Information System for Research in Primary Care (SIDIAP)
4. The Netherlands: Integrated Primary Care Information (IPCI)
5. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Study size

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

Statistical analysis

All analyses will be conducted separately for men and women and will subsequently be stratified by age group, pre-defined comorbidities, and use of hormonal contraceptive/HRT.

Objective 1

Characterisation of initiation of antithrombotic drug treatment following VTE will be conducted. The first ever record of a prescription/dispensation for antithrombotic treatment on the day of VTE diagnosis or within 90 days after will be considered as incident treatment. The number and proportion of individuals who received initial antithrombotic treatments after VTE, both at ingredient and at drug class level, will be reported. Age will be reported as median (interquartile range (IQR)). Baseline comorbidities will be assessed within the year before and on the index date and will be reported as number and proportion.

Objective 2

For the description of the utilisation of antithrombotic treatment following VTE, the index date will be defined as the date of initiation of antithrombotic treatment. Initial dose and initial treatment duration will be described at drug ingredient level. Initial treatment duration will be determined by the duration of the first drug era. Initial dose and initial treatment duration will be reported as median (IQR).

Objective 3

Characterisation of treatment switching patterns in individuals who started antithrombotic treatment will be conducted, with the index date defined as the date of initiation of antithrombotic treatment. Analyses of treatment switching patterns will be conducted separately at ingredient and drug class level. Antithrombotic treatment will be determined on day 0, day 120, and day 240. The number and proportion of individuals receiving each treatment option will be reported at each assessment time point.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

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

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

6. RATIONALE AND BACKGROUND

Antithrombotic agents are used in the management of venous thromboembolism (VTE). Current guidelines recommend direct oral anticoagulants (which inhibit factor Xa or thrombin), (low molecular weight) heparins, and vitamin K antagonists for the prophylaxis of primary and secondary VTE, whilst fibrinolytic agents are used for acute dissolving of clots in Deep Vein Thrombosis (DVT) and Pulmonary Embolism (PE). Clinical guidelines recommend adapted antithrombotic treatments for individuals with renal impairment, cancer, or extreme under/overweight.[1,2]

Optimal treatment may differ between men and women due to biological and clinical factors, including bleeding risk, pharmacokinetics, and hormonal influences, such as oral contraceptives and hormone replacement therapy (HRT). Data from prospective, observational VTE registries indicate that women experience lower treatment satisfaction and poorer health-related quality of life after acute VTE than men [3,4]. Nevertheless, important evidence gaps remain regarding sex-specific antithrombotic treatment strategies, particularly in the management of women receiving hormonal contraception or HRT during antithrombotic therapy.

There is a lack of consistent, large-scale Real-World Evidence to assess whether meaningful treatment gaps exist across European healthcare systems in routine clinical practice. As part of the broader European Medicines Agency (EMA) Women's Health Initiative, this study aims to evaluate potential sex differences in the treatment of VTE. By examining sex-specific patterns in treatment initiation and subsequent treatment pathways, the study seeks to identify potential gaps in care that may inform clinical practice, regulatory decision-making, and public health strategies aimed at promoting equitable, evidence-based treatment.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

To what extent do treatment patterns for thromboembolic disease differ between men and women in clinical practice in European countries?

Objectives

1. To describe antithrombotic treatment initiation for venous thromboembolic events in men and women
2. To describe drug utilisation of antithrombotic treatments, including initial treatment duration and initial dose, in men and women
3. To describe antithrombotic treatment switching patterns for venous thromboembolic events in men and women

All analyses will be conducted overall and stratified by age groups, pre-existing conditions, and use of hormonal contraception/HRT [in women only].

8. RESEARCH METHODS

8.1. Study design

A (new user) cohort study will be conducted using routinely collected health data from 5 data sources from 5 countries across Europe to:

- describe the initiation of antithrombotic drug treatment (at class level [7 classes] and ingredient level (17 pre-specified ingredients) following VTE in men and women (Objective 1).
- describe the initial duration and initial dose of antithrombotics for each of the 17 pre-specified ingredients (Objective 2).
- describe switching patterns of antithrombotics, on a) class level and b) between the 10 most used ingredients (Objective 3).

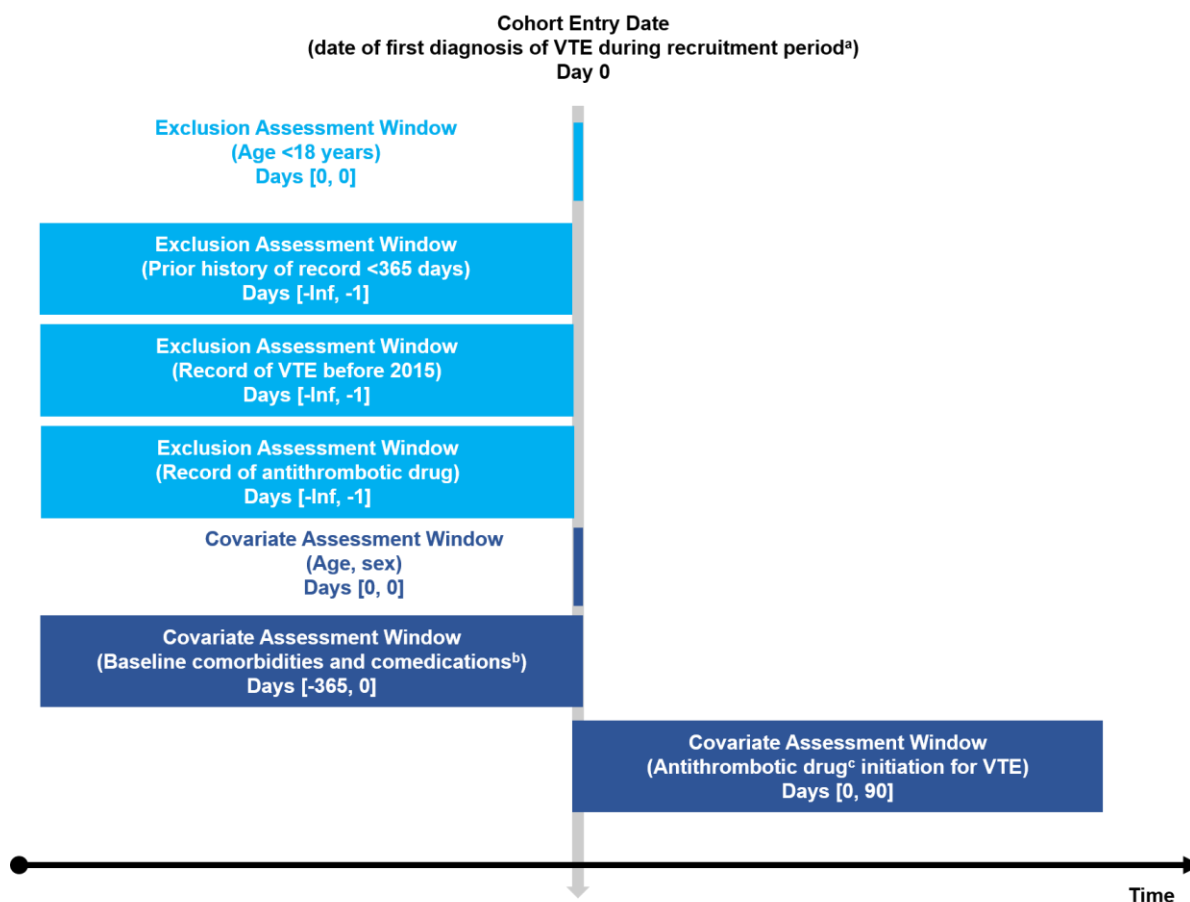


Figure 1. Graphical depiction of the study design of Objective 1.

- a. The recruitment period will be from 01/01/2015 to the earliest of 90 days before the end of the data source observation period or 90 days before end of the study period.
- b. Baseline comorbidities for characterisation/stratification include chronic kidney disease, active cancer, extreme under/overweight; comedications include use of hormonal contraception or hormone replacement therapy.
- c. Antithrombotics will be assessed at ingredient level. Each ingredient or combination regimen will be treated as a mutually exclusive treatment option.

VTE = venous thromboembolic event

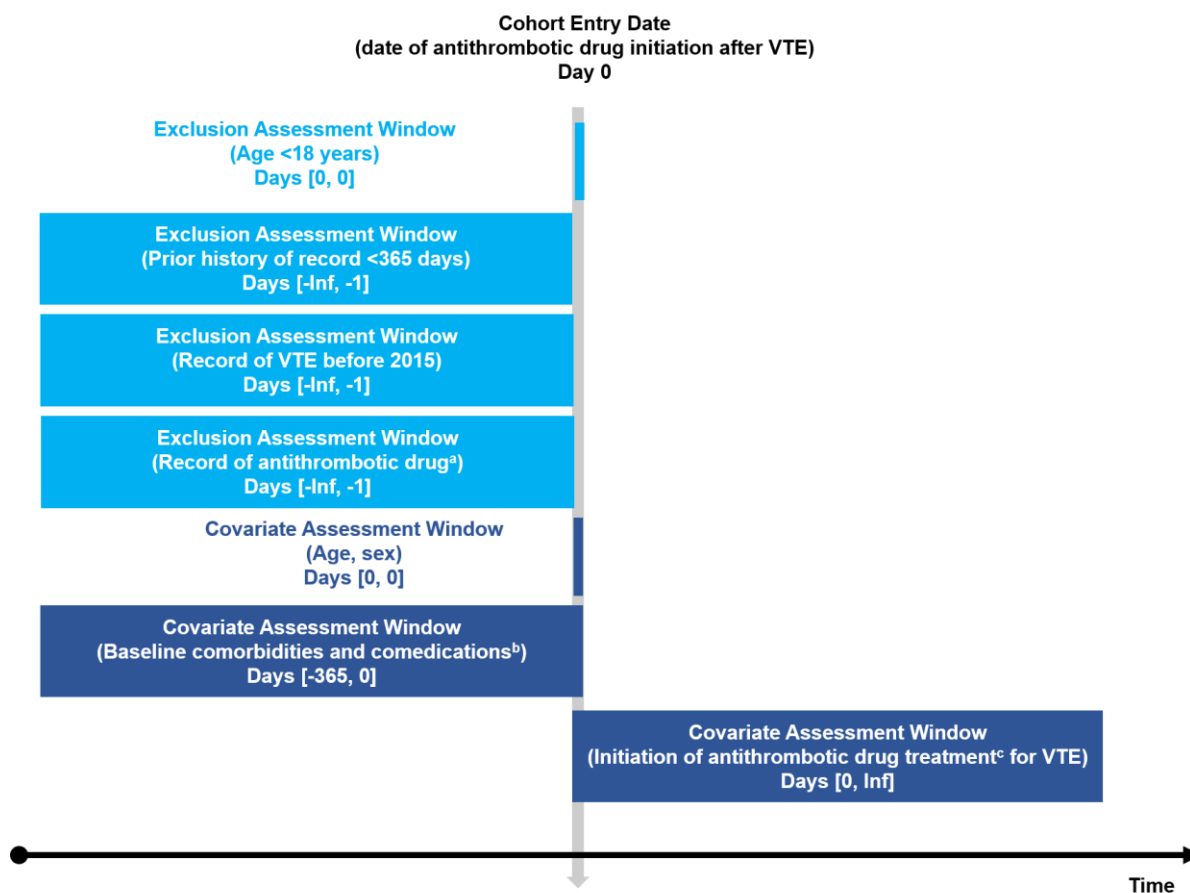


Figure 2. Graphical depiction of the study design of Objective 2.

- a. Antithrombotics will be assessed at ingredient level.
- b. Baseline comorbidities for characterisation/stratification include chronic kidney disease, active cancer, extreme under/overweight; comedications include use of hormonal contraception or hormone replacement therapy.
- c. Drug utilisation of treatment duration of first drug era (with gap era of 90 days) and initial dose of first drug exposure record will be assessed.

VTE = venous thromboembolic event

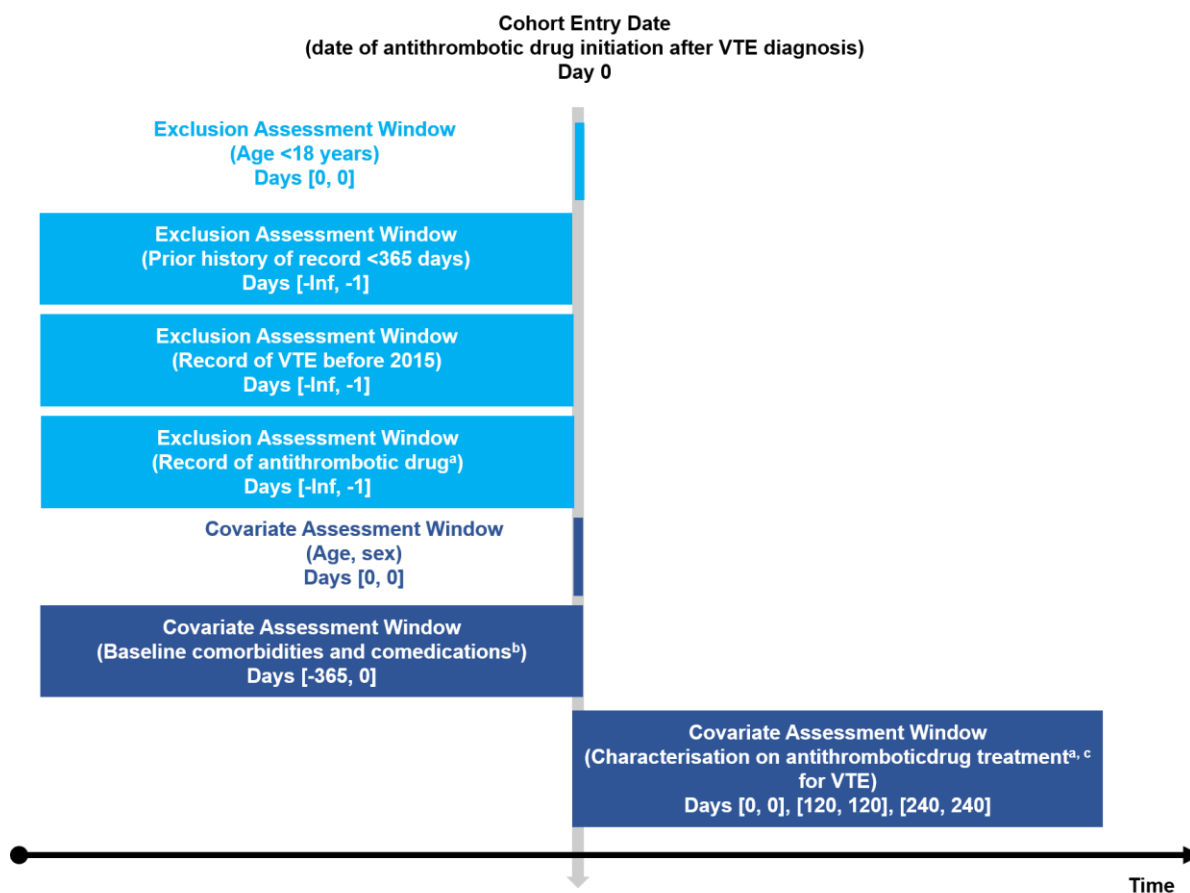


Figure 3. Graphical depiction of the study design of Objective 3.

- a. Antithrombotics will be assessed at ingredient level.
- b. Baseline comorbidities for characterisation/stratification include chronic kidney disease, active cancer, extreme under/overweight; comedications include use of hormonal contraception or hormone replacement therapy.
- c. Treatment patterns of antithrombotics will be assessed at ingredient level. Each ingredient or combination regimen will be treated as a mutually exclusive treatment option.

VTE = venous thromboembolic event

8.2. Follow-up

Objective 1: Individuals with a first event of VTE (deep vein thrombosis (DVT) or pulmonary embolism (PE)) during the recruitment period will be included. Follow-up will start on the date of first VTE record (index date). Follow-up will end on the earliest of loss to follow-up, death, end of the observation period (the latest available data), study end date (31/12/2025), or 90 days after the index date. A 90-day window has been chosen, as it is expected that antithrombotics are initiated immediately following VTE. However, in case a first hospital prescription would be missed, the window would allow capture of the subsequent outpatient prescription.

Objective 2: For analyses describing initial dose and duration, follow-up will start on the date of antithrombotic treatment initiation (index date). Follow-up will end on the earliest of loss to follow-up, death, end of the observation period (the latest available data), or the study end date (31/12/2025).

Objective 3: For the characterisation of antithrombotic treatment switching patterns, follow-up will start on the date of antithrombotic drug treatment initiation (index date). Follow-up will end on the earliest of loss to follow-up, death, end of the observation period (the latest available data), the study end date

(31/12/2025), or 240 days after the index date. This window of follow-up was defined to capture any potential treatment switching pattern in individuals requiring prolonged anticoagulation treatment, as recommended by the NICE clinical guideline.

8.3. Study population with inclusion and exclusion criteria

Objective 1 (Characterisation of antithrombotic treatment initiation)

Inclusion criterion: First recorded diagnosis of VTE (DVT or PE) during the recruitment period

Exclusion criteria

- Age <18 years at first recorded diagnosis of VTE (index date)
- Less than 365 days of available history before the index date
- Record of VTE before study start (01/01/2015)
- Record of antithrombotic drug prescription/dispensation before the index date

Objectives 2 and 3 (Drug utilisation of antithrombotic treatment and characterisation of treatment switching patterns)

Inclusion criteria

- First recorded diagnosis of VTE during the recruitment period
- Initiation of antithrombotic treatment within 90 days after diagnosis (index date)

Exclusion criteria

- Age <18 years on the date of antithrombotic treatment initiation
- Less than 365 days of available history before date of antithrombotic treatment initiation
- Record of VTE before study start 01/01/2015
- Record of antithrombotic drug prescription/dispensation before first VTE

8.4. Study setting and data sources

This study will be conducted using routinely collected data from five primary and secondary care data sources in the DARWIN EU® Network of Data Partners from five European countries. All data were *a priori* mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (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
Denmark	Danish Data Health Registries (DK-DHR)	Community pharmacies, secondary care – specialists, hospital inpatient care	Registries	6.0M	1st January 2015 to 31st December 2025
Finland	Finnish Care Register for Health Care (FinOMOP-THL)	Primary care, secondary care – specialists (ambulatory or hospital OP	EHR, Registries	6.2M	1st January 2015 to 31st December 2025

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source
		care), hospital IP care			
Spain	The Information System for Research in Primary Care (SIDIAP)	Primary care data source + linkage to hospital data	EHR	6.1M	1st January 2015 to 31st December 2024
The Netherlands	Integrated Primary Care Information (IPCI)	Primary care	EHR	1.3M	1st January 2015 to 31st December 2025
The United Kingdom	Clinical Practice Research Datalink (CPRD GOLD)	Primary care	EHR	2.4M	1st January 2015 to 1st December 2025

EHR = electronic health record; IP = inpatient; M = million; OP = outpatient.

Data sources selection

These data sources fulfil the criteria required in terms of data quality, completeness, timeliness, availability to conduct the study within the required timelines, and representativeness for patient-level drug utilisation study while covering different regions of Europe ([Annex II](#)).

8.5. Study period

The study period is from 01/01/2015 to 31/12/2025 or the date of the most recent data available for each contributing data source.

The study recruitment period is from 01/01/2015 to the earliest of 90 days before the end of the data source observation period or 30/9/2025 (90 days before the end of the study period).

8.6. Variables

8.6.1. Exposure

Diagnosis of VTE

All individuals with a (first ever) diagnosis of VTE recorded during the study period will be selected. VTE will be defined as diagnosis of DVT or PE.

The computational phenotypes for VTE have been developed during another DARWIN EU® study ([P3-C3-001](#)) and are available in the DARWIN EU® library of phenotypes.

Antithrombotic drugs and drug classes

Exposure to antithrombotic treatment will be identified as the first recorded prescription or dispensing for antithrombotic drugs after a first record of VTE. For DK-DHR, CPRD GOLD, IPCI, and SIDIAP, prescription records will be used, whereas for FinOMOP-THL, dispensation records will be used.

For each ingredient, treatment episodes of sequential prescriptions (i.e. drug era) of the same ingredient will be generated. A maximum of 90 days between the end date of one prescription/dispensation record

and the start date of the next prescription/dispensation record (i.e. definition of gap) will be allowed for them to be combined to continuous treatment eras.

By concatenating individual drug exposure records into drug eras, the exposed risk window will account for all periods during which the individual is likely to be using the respective ingredients.

The preliminary concept sets used for the identification of exposures are described in [Annex IV](#).

Drug classes/groups under Anatomical Therapeutic Chemical [classification] (ATC) code B01 (antithrombotic agents) include the following: B01AA – Vitamin K antagonist, B01AB – Heparin, B01AC – Platelet aggregation inhibitors, B01AD – Enzymes, B01AE – Direct thrombin inhibitors, B01AF – Direct factor Xa inhibitors, B01AX – Other antithrombotic agents). Ingredients belonging to these classes/groups and prescribed in the 5 data sources as well as the corresponding feasibility counts are provided in [Annex V](#).

Clinical guidelines for VTE treatment recommend direct factor Xa and thrombin inhibitors, (low molecular weight) heparins, and vitamin K antagonists. Fibrinolytic enzymes might be used for the acute dissolving of clots in DVT and PE. Therefore, the following pre-specified ingredients have been prioritised for inclusion in the ingredient-level analyses in the respective objectives of this study ([Table 2](#)). Ingredient-level analyses will exclude ingredients belonging to platelet aggregation inhibitors. Drug class level analyses will include all seven ATC groups ([Table 3](#)).

Table 2. Pre-specified antithrombotics for ingredient-level analyses.

ATC Class	Ingredient name	Objective 1	Objective 2	Objective 3
B01AA	acenocoumarol	X	X	X
B01AD	alteplase	X	X	
B01AF	apixaban	X	X	X
B01AB	bemiparin	X	X	
B01AE	dabigatran	X	X	X
B01AB	dalteparin	X	X	X
B01AF	edoxaban	X	X	X
B01AB	enoxaparin	X	X	X
B01AX	fondaparinux	X	X	
B01AB	heparin	X	X	
B01AB	nadroparin	X	X	
B01AA	phenprocoumon	X	X	X
B01AF	rivaroxaban	X	X	X
B01AD	streptokinase	X	X	
B01AB	sulodexide	X	X	
B01AB	tinzaparin	X	X	X
B01AA	warfarin	X	X	X

ATC = Anatomical Therapeutic Chemical [classification]

Table 3. Pre-specified antithrombotics for drug class-level analyses.

ATC Class	Class Name	Objective 1	Objective 2	Objective 3
B01AA	Vitamin K antagonist	X	X	X
B01AB	Heparin	X	X	X
B01AC	Platelet aggregation inhibitors	X	X	X
B01AD	Enzymes	X	X	
B01AE	Direct thrombin inhibitors	X	X	X
B01AF	Direct factor Xa inhibitors	X	X	X
B01AX	Other antithrombotic agents	X	X	

ATC = Anatomical Therapeutic Chemical [classification]

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

All analyses will be conducted separately for men and women.

Objective 1

The following covariates will be characterised on the date of first recorded diagnosis of VTE (index date) and used for stratification (i.e. age groups, pre-defined comorbidities, use of hormonal contraceptives, use of hormone replacement therapy)

- Age at the index date (median, interquartile range (IQR))
- Age groups at the index date (N, %)
 - 18 – 29 years
 - 30 – 49 years
 - 50 – 64 years
 - 65+ years
- Comorbidities, with assessment window [-365, 0] days relative to the index date (N, %)
 - chronic kidney disease
 - active cancer (based on diagnosis in the last year or prescription of antineoplastic/chemotherapeutic treatment in the last year)
 - extreme under/overweight (based on diagnosis, or Body Mass Index (BMI) measurement $\leq 18.5 \text{ kg/m}^2$ or $\geq 30 \text{ kg/m}^2$, respectively)
- Comedication, with assessment window [-365, 0] days relative to the index date (N, %)
 - Use of hormonal contraceptives (yes/no)
 - Use of hormone replacement therapy (yes/no)

Objectives 2 and 3

The following covariates, identified on the date of initiation of antithrombotic treatment (index date), will be used for stratification:

- Age groups at the index date
 - 18 – 29 years
 - 30 – 49 years
 - 50 – 64 years
 - 65+ years
- Comorbidities, with assessment window [-365, 0] days relative to the index date
 - chronic kidney disease
 - active cancer (based on diagnosis in the last year or prescription of antineoplastic/chemotherapeutic treatment in the last year)
 - extreme under/overweight (based on diagnosis, or BMI measurement $\leq 18.5 \text{ kg/m}^2$ or $\geq 30 \text{ kg/m}^2$, respectively)
- Comedication, with assessment window [-365, 0] days relative to the index date (N, %)
 - Use of hormonal contraceptives (yes/no)
 - Use of hormone replacement therapy (yes/no)

8.7. Study size

No sample size has been calculated, as this study will not test a specific hypothesis. In addition, already collected available data will be used to estimate patient characteristics, treatment characteristics, and treatment switching patterns. Thus, the sample size is driven by the availability of data for individuals with exposure to antithrombotic drugs. Complete feasibility counts for included antithrombotic drugs are included in [Annex V](#).

8.8. Analysis

8.8.1. Federated network analyses

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

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

8.8.2. Data privacy protection

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

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

All analyses will be conducted separately for men and women and will subsequently be stratified by age group, pre-defined comorbidities, and use of hormonal contraceptives/use of hormone replacement therapy [in women only] ([Section 8.6.2](#)).

Objective 1

Characterisation of antithrombotic drug treatment following VTE will be conducted based on OMOP CDM mapped data using the R package *CohortCharacteristics*,[5] developed by DARWIN EU®.

The index date will be defined as the date of first diagnosis of VTE. Incident treatment will be assessed with the assessment window [0, 90] days relative to the index date. The first record of a prescription/dispensation for antithrombotic treatment after diagnosis of VTE will be considered as initial treatment. Combinations of multiple ingredients, e.g. the prescription or dispensation of different drugs on the same date, which is the date of antithrombotic treatment initiation, will be allowed. Monotherapies (e.g. B01AF) and combinations of ingredients (e.g. B01AF + B01AC) will be reported as mutually exclusive groups, i.e. an individual with a first prescription of B01AF + B01AC will only be included in the B01AF + B01AC group, and not in the monotherapy groups for B01AF and B01AC, respectively.

The number and proportion of individuals who initiated antithrombotic treatments after VTE, both at ingredient (17 pre-specified ingredients) and at drug class level (single or combination, mutually exclusive groups), will be reported. The median (IQR) time between index date (incident VTE diagnosis) and treatment initiation will be reported. Age will be reported as median (IQR). Baseline comorbidities will be assessed with the assessment window [-365, 0] relative to the index date and will be reported as number and proportion.

Objective 2

Drug utilisation of antithrombotic treatment following VTE will be conducted based on OMOP CDM mapped data using the R package *DrugUtilisation*,[6] developed by DARWIN EU®.

The index date will be defined as the date of initiation of antithrombotic treatment. Initial dose and initial treatment duration will be described at drug ingredient level for the 17 pre-specified ingredients and for monotherapy only.

The initial dose will be determined by the daily dose prescribed/dispensed on the index date. Treatment episodes of sequential prescriptions (i.e. drug era) of the same drug ingredient that have a maximum of 90 days between the end date of one prescription and the start date of the next prescription (i.e. definition of gap), will be generated. Initial treatment duration will be determined by the duration of the first drug era. Initial dose and initial treatment duration will be reported as median (IQR).

Objective 3

Characterisation of treatment switching patterns in individuals who started antithrombotic treatment will be conducted based on OMOP CDM mapped data using the R package *PatientProfiles*,[7] developed by DARWIN EU®.

The index date will be the date of antithrombotic treatment initiation. Treatment switching pattern analyses will be conducted separately at ingredient level (10 pre-specified ingredients ([Table 2](#), column “Objective 3”)) and on the drug class level (7 drug classes). Drug eras will be computed, allowing for 90 days between subsequent prescription of ingredients of the same drug class. As per guidance, initial anticoagulation therapies are recommended during the first 3 months after VTE (with 3–6 months longer duration for cancer patients). Subsequently, long-term anticoagulation for secondary prevention might be needed. Therefore, in this study, antithrombotics will be determined on day 0 (index date), day 120 (to capture potential treatment stop/switching at 3 months), and day 240 (to capture potential treatment

stop/switching at 6 months). Similar to Objective 1, combinations of ingredients and ingredients belonging to different antithrombotic drug classes as treatment on the index date will be allowed. The number and proportion of individuals receiving each treatment option will be reported at each assessment time point. Sankey diagrams will be used to illustrate switching patterns.

8.8.4. Output

Output will include the following:

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

- Table 1. Characteristics of baseline comorbidities and initial treatment of individuals with VTE, in men and women.
- Table 2. Characteristics of baseline comorbidities and initial treatment of individuals with VTE, in men and women, stratified by age group.
- Table 3. Drug utilisation of initial treatment for VTE, in men and women.
- Figure 1. Sankey plot for treatment switching patterns at drug-class level in men, per data source.
- Figure 2. Sankey plot for treatment switching patterns at drug-class level in women, per data source.
- Figure 3. Sankey plot for treatment switching patterns at ingredient level in men, per data source.
- Figure 4. Sankey plot for treatment switching patterns at ingredient level in women, per data source.

Mock tables and figures are available in [Annex VI](#).

Analyses stratified by baseline comorbidities will be available in the Shiny app. An interactive dashboard will be generated by incorporating all the results (tables and figures) included in the PDF report mentioned above.

8.9. 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.

9. STRENGTHS AND LIMITATIONS

The study will be informed by routinely collected health care data and so, data quality issues must be considered. While antithrombotics prescribed or administered in primary care or outpatient settings are likely well recorded, certain classes of antithrombotics, namely fibrinolytics, used in hospital as emergency medicines to dissolve blood clots are not recorded in most data sources.

Prescriptions and/or purchase of hormonal contraception differs between countries: hormonal contraceptives are only available through prescription in Denmark, while in the UK, over-the-counter progestogen-only contraceptive pills are available since 2021. Contraceptives can also be prescribed in Sexual Health Clinics; therefore, it might not be possible to capture this information in general practitioner data. In Spain, only specific types of hormonal contraceptives are included in the universal healthcare, whilst others can be purchased over the counter.

Moreover, the documentation of comorbidities, necessary for characterisation and stratification, may vary across data sources. Likewise, the identification of thromboembolic events may vary across data sources. As the analyses will be restricted to individuals with a recorded VTE, relatively few false positive diagnoses

would be expected, but true events could potentially be missed, especially in primary care data sources without patient-level linkage to secondary care data.

While the study aims to describe potential differences in treatment patterns between men and women, no matching has been performed, and differences in treatment might not be solely attributable to sex but other factors, such as comorbidities, including those related to bleeding risk. For instance, a study based on a multinational VTE registry reported fewer DVT recurrences and more bleeds in women compared to men during the course of antithrombotic treatment. However, these differences may not be solely attributed to the difference in sex, but likely due to other patient characteristics more common in women and differences in the choice of treatment.[8]

Additionally, the results estimated from this study will only reflect the populations from the included data sources. Electronic health records have certain inherent limitations because they were collected for clinical purposes rather than primarily for research use.

10. REFERENCES

- [1]. NICE Guideline NG158: Venous thromboembolic diseases: diagnosis, management and thrombophilia testing. Accessed 29/06/2026 via <https://www.nice.org.uk/guidance/NG158> and <https://www.nice.org.uk/guidance/ng158/resources/visual-summary-pdf-11193380893>
- [2]. Konstantinides SV, Meyer G, Becattini C, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS): The Task Force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). *European Heart Journal* 2020; **41**(4): 543–603.
- [3]. Giustozzi M, Valerio L, Agnelli G et al. Sex-specific differences in the presentation, clinical course, and quality of life of patients with acute venous thromboembolism according to baseline risk factors. Insights from the PREFER in VTE. *European Journal of Internal Medicine*, 2021; 88, 43-51.
- [4]. Prandoni P, Fluharty M, Schellong S et al. Sex differences in venous thromboembolism outcomes: findings from the GARFIELD-VTE registry. *European Journal of Internal Medicine*, 2025; 144
- [5]. Du M, Prats-Urbe A, Mercadé-Besora N, et al. CohortCharacteristics: an R package for population characterisation in observational studies using the OMOP common data model. *Eur J Epidemiol*. 2026.
- [6]. Catala M, Du M, Guo Y, et al. DrugUtilisation: Summarise Patient–Level Drug Utilisation in Data Mapped to the OMOP Common Data Model [Available from: <https://cran.r-project.org/web/packages/DrugUtilisation/index.html>].
- [7]. Catala M, Guo Y, Du M et al: PatientProfiles: Identify Characteristics of Patients in the OMOP Common Data Model [Available from: <https://cran.r-project.org/web/packages/PatientProfiles/index.html>]
- [8]. Blanco-Molina A, Enea I, Gadelha T, et al. Sex differences in patients receiving anticoagulant therapy for venous thromboembolism. *Medicine (Baltimore)*. 2014 Oct;93(17):309-317

11. ANNEXES

ANNEX I. Description of data sources

Denmark: Danish Data Health Registries (DK-DHR)

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

#	Section	Description
12	Limitations	<ul style="list-style-type: none"> - Lifestyle factors (such as smoking status and weight) are only complete for pregnant women. - Stillborn children will not have any records in our CDM. This means that e.g. birth weight and birth length of stillborns is not recorded. - There is no information on key socio-economic status (SES) factors, such as occupation, education, and income. - Patient contacts in primary care (visits to the GP) are not included. Consequently, the incidence of chronic diseases like Type 2 Diabetes (T2D) and asthma must be determined using drug prescriptions as a proxy.
13	Main references	Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V, Sørensen HT "The Danish health care system and epidemiological research: from health care contacts to database records." Clinical epidemiology (2019): 31372058
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111217 Website: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/healthdatadenmark

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

#	Section	Description
1	Data source identification and country	FinOMOP-THL (Finnish Care Register for Health Care) Finland
2	Data partner information section	FinOMOP (FinOMOP) The Department of Data and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. The current CDM population comprises all persons having been alive and residing in Finland since the beginning of 2011.
4	Healthcare setting / type of data	<p>Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify).</p> <p>THL maintains health registers that cover both public and private, primary, and specialised inpatient, urgent and outpatient health care encounters in Finland, starting from 2011. The entire public sector and private inpatient encounters have been included since 2011, while private outpatient encounters, including occupational care, are included since 2020.</p> <p>Since 1998, the Care Register for Health Care (TerveysHilmo) has covered both public outpatient and inpatient specialized care and private inpatient care. Since 2009, the Finnish National Vaccination Register has covered all vaccinations from the public sector and from a large part of private vaccination providers, with the data coverage from both sections being very good to complete from 2020 onwards.</p> <p>Since 2011, the Register of Primary Health Care Visits (AvoHilmo) has covered public primary care. Since 2020, the register has also covered private outpatient care and occupational care.</p> <p>In addition, the CDM also contains positive COVID-19 test results from the Finnish National Infectious Diseases Register. The register itself covers all laboratory confirmations for around 70 specific microbes from 1995 onwards, but only COVID-19 has currently been mapped to the CDM.</p> <p>The CDM also includes prescription records from multiple different sources. Both Care Register for Health Care and Register of Primary Health Care Visits contain very basic prescription data recorded during health care encounters that include just the ATC-code and trade name of medication. More comprehensive prescription data from Kanta Prescription Centre, maintained by Social Insurance Institution of Finland (Kela), has been integrated into the CDM since its 2024 release. The Kanta Prescription records are based electronic prescriptions, which were adopted by most public health care providers in 2010 and by most private providers by 2017.</p>
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. Data is entered by clinicians upon healthcare contact and processed by THL (Kela in the case of Kanta Prescription Centre).

#	Section	Description
6	General representativeness	The THL data has national coverage and is therefore well representative of the Finnish population. Using the complete population as a basis for the person table also serves to facilitate calculations on a population level, e.g. incidence rates.
7	Data content / source coding	The following coding systems have been OMOP-mapped, typically to a good level of completeness: ICD10fi Finnish Extension, ICPC-2, ATC, Toimenpideluokitus (procedure classification adapted from the Nordic Classification of Surgical Procedures (NCSP)), Terveystieteiden tutkimuskeskusten erikoissalat (Hilmo specific provider speciality), Rokotustapa (AR/YDIN National classification for vaccine administration), Tupakointistatus (AR/YDIN National classification for smoking status). Vaccinations are identified on product level based on batch number, trade name, vaccine title, and ATC-code. This is mapped on brand and type in the OMOP CDM.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Each patient in THL has a unique identifier.
9	Quality control (data source specific)	The source data collection undergoes a structural and semantic validation before entry into the source database. Additionally, some coded variables undergo quality assessment against the respective code systems post entry into the database. The source registers are also assessed for completeness and coverage, with the aim of improving future collection in the areas where data is lacking.
10	Linkage	THL is already a linkage of multiple Finnish registries (see above).
11	Vital status	The National Population registry data forms the basis for forming the patient population. This ensures an up-to-date location (municipality of residence) of patients, as well as complete death occurrences (although not the cause of death).
12	Limitations	All drug records in the CDM are currently based on prescriptions. Kanta Prescription Centre also includes information on drug dispensings, but these have not currently been converted into OMOP CDM. Depending on the type of the medication, a single prescription can be for up to two years dosage of the drug. This means a patient can have up to two-year breaks between observations while actively using the drug. Observation of a prescription also does not mean that the patient necessarily bought or used the medication. The CDM does not currently have meaningful end dates or days of supply for drug exposure. This information is not available for Care Register for Health Care or Register of Primary Health Care Visits, and for Kanta Prescription Centre it is only available in unstructured, free-form format that has not been converted into OMOP CDM. The source system for prescription drug data is only available to the DP until Dec-2023, therefore drug_exposure records from this source are only available in that time range. Some vital signs, like height and weight measurements, are recorded even if the person has not been actually measured, just that the measurement is deemed relevant to the condition or it's treatment. This may lead to duplicate measurements as the value from previous visit might be taken. Not all private health care records are covered for the CDM's entire follow-up time from 2011 onwards. For Register of Primary Health Care Visits and Finnish National Vaccination Register, records from private health care have been available from 2020 onwards. For Kanta Prescription Centre, the coverage of private health care records has been good from 2017 onwards. The inclusion of private health care mainly presents itself as an increase in the number of observations, meaning that it has to be accounted for when interpreting any time series data from the CDM.
13	Main references	Häkkinen, Pirjo; Mölläri, Kaisa; Saukkonen, Sanna-Mari; Väyrynen, Riikka; Mielikäinen, Lasse; Järvelin, Jutta "Hilmo - Sosiaali- ja terveydenhuollon hoitoilmoitus 2020 : Määrittelyt ja ohjeistus : Voimassa 1.1.2020 alkaen" Terveystieteiden tutkimuskeskusten erikoissalat (2019):
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111187 ; https://catalogues.ema.europa.eu/data-source/1111191 Website: https://thl.fi/en/statistics-and-data/data-and-services/register-descriptions ; https://www.kanta.fi/en/research-and-knowledge-management

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

#	Section	Description
1	Data source identification and country	SIDIAP (The Information System for Research in Primary Care) Catalunya, SB01ACn
2	Data partner information section	Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. SIDIAP is a database of primary care electronic health records of the population of Catalonia, North-East SB01ACn. It contains pseudo-anonymised records of more than 8 million individuals , of which 6.1 million are active as of 2022, representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – General Practitioner, and hospital inpatient care. SIDIAP captured data includes routine visits, socio-demographics, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, sick leaves and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Català de la Salut (Catalan Health Institute) is the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content / source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient has a unique id, also upon changing practice or re-registration.
9	Quality control (data source specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g., there are several variables with information concerning the 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.

#	Section	Description
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, Bolívar B, Aragón M, Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

The Netherlands: Integrated Primary Care Information (IPCI)

#	Section	Description
1	Data source identification and country	IPCI (Integrated Primary Care Information) the Netherlands
2	Data partner information section	Erasmus University Medical Center (EMC) Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards. However, it mainly covers the central part of the country, including the most densely populated area (the 'Randstad') and non-urban areas. IPCI contains information on all patients registered with GPs responsible for non-emergency care and referrals. A patient is registered at birth or at first encounter with the GP.
4	Healthcare setting / type of data	Primary care – General Practitioner. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. The patientdossiers are collected by Erasmus MC data managers and combined in one harmonized database. Several checks are done on this database to ensure correct data processing. Persons can have dossiers at multiple GPs.
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content / source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients can be registered under different IDs, but since a patient can only be registered at one GP at a time, the observations periods will not overlap.

#	Section	Description
9	Quality control (data source specific)	<p>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.</p> <p>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.</p>
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	<p>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.</p> <p>Capture of oral contraceptives in IPCI is incomplete. The first prescription is recorded within the patient's medical file but not the subsequent prescriptions. If you want to use IPCI data for research questions on oral contraceptives, you can do that but then you need to select those GPs that use MEDICOM as software system (where you have prescribing and dispensing). However, of hormone substitution therapy, the information on prescribing is COMPLETE. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.</p>
13	Main references	de Ridder MAJ, de Wilde M, de Ben C, Leyba AR, Mosseveld BMT, Verhamme KMC, van der Lei J, Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182144
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

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

#	Section	Description
1	Data source identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) the United Kingdom
2	Data partner information section	University of Oxford - Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences (Oxford NDORMS) NDORMS
3	Coverage and timespan	<p>Data collection since: 1987</p> <p>Extent: Nation-wide.</p> <p>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,</p>

#	Section	Description
		no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records. Data are entered by clinicians into the EHR. Data is processed by CPRD that provides data releases for research.
6	General representativeness	In the last 10 years, the CPRD GOLD regional distribution of currently contributing general practitioner (GP) practices has significantly shifted, resulting in many new practices joining from Scotland, Wales, and Northern Ireland, and fewer participating from England. These changes have affected the CPRD GOLD population size, regional coverage, and eligibility for data linkages. CPRD GOLD January 2024 contains >21.3 million historical and current patients (12.9 in England, 3.1 in Wales, 4.7 in Scotland, 0.7 in Northern Ireland). Of these, nearly 3 million are currently registered in a GP practice and represent ~4.3% of the estimated current UK population (0.1% in England, 32.3% in Wales, 28.6% in Scotland, 16.2% in Northern Ireland). Patients currently registered in CPRD GOLD January 2024 are broadly representative of the UK population with respect to age and sex. Reference: https://doi.org/10.1093/ije/dyaf077
7	Data content / source coding	Gemscript, Read, dm+d
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In GOLD, a patient can be registered under different ID numbers upon changing practice or re-registration. Researchers are not able to identify these patients, as the data are anonymised. However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (data source specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (UTS). Each practice is associated to an UTS date set when the data quality standards become satisfactory, and CPRD recommend using only longitudinal data starting from this UTS date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards, and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice UTS date. CPRD also checks data quality standards at the patient level, and associates each patient with a flag, reporting if its data are acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at the Practice Level.
11	Vital status	The date of death in CPRD GOLD has been validated against the Population registry (ONS) mortality data. Reference: https://doi.org/10.1002/pds.4747
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL, Dedman D, Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193

#	Section	Description
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://www.cprd.com/data/primary-care-data/cprd-gold

ANNEX II. Fitness for use assessment

Danish Data Health Registries (DK-DHR)

DK-DHR will be included in this study because it is a registry data source that provides relevant information on prescription records of antithrombotics in the general adult population.

Moreover, data availability and follow-up in DK-DHR is sufficient, as data availability starts in 1995, and the date of the most recent data extraction is 4th June 2026, which aligns with the study period.

There are no study specific limitations present in DK-DHR.

Lastly, DK-DHR has blanket Institutional Review Board (IRB) approval, which makes the execution of this study feasible within the current study timelines.

Finnish Care Register for Health Care (FinOMOP-THL)

FinOMOP-THL will be included in this study because it is a primary and secondary care data source that provides relevant information on prescription records of antithrombotics in the general adult population.

Moreover, data availability and follow-up in FinOMOP-THL is sufficient, as data availability starts in 1998, and the date of the most recent data extraction is 14th April 2026, which aligns with the study period.

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

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

The Information System for Research in Primary Care (SIDIAP)

SIDIAP will be included in this study because it is a primary and secondary care data source that provides relevant information on prescription records of antithrombotics in the general adult population.

Moreover, data availability and follow-up in SIDIAP is sufficient, as data availability starts in 2006, and the date of the most recent data extraction is 31st December 2024, which aligns with the study period.

There are some study specific limitations present in SIDIAP, namely absence of data from December 31st 2024 onwards.

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

Integrated Primary Care Information (IPCI)

IPCI will be included in this study because it is a primary care data source that provides relevant information on prescription records of antithrombotics in the general adult population.

Moreover, data availability and follow-up in IPCI is sufficient, as data availability starts in 2006, and the date of the most recent data extraction is 19th April 2026, which aligns with the study period.

There are no study specific limitations present in IPCI.

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

Clinical Practice Research Datalink GOLD (CPRD GOLD)

CPRD GOLD will be included in this study because it is a primary care data source that provides relevant information on prescription records of antithrombotics in the general adult population.

Moreover, data availability and follow-up in CPRD GOLD is sufficient, as data availability starts in 1987, and the date of the most recent data extraction is 1st December 2025, which aligns with the study period.

There are some study specific limitations present in CPRD GOLD, namely absence of data from 1st December 2025 onwards.

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

ANNEX III. Operational and reporting considerations

DATA MANAGEMENT

Data management

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

The analytic code for this study will be written in R and will use standardised analytics wherever possible. Each data partner will execute the study code against their data source containing patient-level data and then return the results (csv files), which will only contain aggregated data. The results from each of the contributing data sites will then be combined in tables and figures for the study report.

Data storage and protection

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

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

The output files are stored in the DARWIN EU® Digital Research Environment (DRE). These output files do not contain any data that allow identification of subjects included in the study. The DRE implements further security measures to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/2016 in the various member states.

QUALITY CONTROL

Data source quality control

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

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

The study code will be based on DARWIN EU® R packages: *DrugUtilisation* to characterise the drug use, and *CohortCharacteristics* and *PatientProfiles* 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 EMA by the DARWIN EU® CC upon completion of the study.

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

ANNEX IV. List of concepts

Table S1. Preliminary list of condition concept IDs for defining VTE.

Concept ID	Concept name	domain	vocabulary
Deep vein thrombosis			
762047	Acute bilateral thrombosis of subclavian veins	Condition	SNOMED
762148	Acute deep vein thrombosis of bilateral iliac veins	Condition	SNOMED
35616028	Acute deep vein thrombosis of left iliac vein	Condition	SNOMED
35615035	Acute deep vein thrombosis of left lower limb following procedure	Condition	SNOMED
35615031	Acute deep vein thrombosis of left upper limb following procedure	Condition	SNOMED
43531681	Acute deep vein thrombosis of lower limb	Condition	SNOMED
35616027	Acute deep vein thrombosis of right iliac vein	Condition	SNOMED
44782746	Acute deep venous thrombosis	Condition	SNOMED
44782751	Acute deep venous thrombosis of axillary vein	Condition	SNOMED
762008	Acute deep venous thrombosis of bilateral axillary veins	Condition	SNOMED
760875	Acute deep venous thrombosis of bilateral calves	Condition	SNOMED
765155	Acute deep venous thrombosis of bilateral iliofemoral veins	Condition	SNOMED
762017	Acute deep venous thrombosis of bilateral internal jugular veins	Condition	SNOMED
762417	Acute deep venous thrombosis of bilateral legs	Condition	SNOMED
762020	Acute deep venous thrombosis of bilateral popliteal veins	Condition	SNOMED
765546	Acute deep venous thrombosis of bilateral tibial veins	Condition	SNOMED
762004	Acute deep venous thrombosis of both upper extremities	Condition	SNOMED
44782742	Acute deep venous thrombosis of calf	Condition	SNOMED
44782747	Acute deep venous thrombosis of femoral vein	Condition	SNOMED
762015	Acute deep venous thrombosis of iliofemoral vein of left leg	Condition	SNOMED
765541	Acute deep venous thrombosis of iliofemoral vein of right lower extremity	Condition	SNOMED
44782748	Acute deep venous thrombosis of iliofemoral vein	Condition	SNOMED
44782752	Acute deep venous thrombosis of internal jugular vein	Condition	SNOMED
762009	Acute deep venous thrombosis of left axillary vein	Condition	SNOMED
760876	Acute deep venous thrombosis of left calf	Condition	SNOMED
765540	Acute deep venous thrombosis of left femoral vein	Condition	SNOMED
765922	Acute deep venous thrombosis of left internal jugular vein	Condition	SNOMED
762418	Acute deep venous thrombosis of left lower extremity	Condition	SNOMED
765537	Acute deep venous thrombosis of left upper extremity	Condition	SNOMED
762022	Acute deep venous thrombosis of popliteal vein of right leg	Condition	SNOMED
44782743	Acute deep venous thrombosis of popliteal vein	Condition	SNOMED
762021	Acute deep venous thrombosis of popliteal vein of left leg	Condition	SNOMED
762010	Acute deep venous thrombosis of right axillary vein	Condition	SNOMED
760877	Acute deep venous thrombosis of right calf	Condition	SNOMED

Concept ID	Concept name	domain	vocabulary
762013	Acute deep venous thrombosis of right femoral vein	Condition	SNOMED
762018	Acute deep venous thrombosis of right internal jugular vein	Condition	SNOMED
762419	Acute deep venous thrombosis of right lower extremity	Condition	SNOMED
762005	Acute deep venous thrombosis of right upper extremity	Condition	SNOMED
44782745	Acute deep venous thrombosis of thigh	Condition	SNOMED
44782744	Acute deep venous thrombosis of tibial vein	Condition	SNOMED
762026	Acute deep venous thrombosis of tibial vein of left leg	Condition	SNOMED
765156	Acute deep venous thrombosis of tibial vein of right leg	Condition	SNOMED
44782421	Acute deep venous thrombosis of upper extremity	Condition	SNOMED
762048	Acute thrombosis of left subclavian vein	Condition	SNOMED
45757410	Acute thrombosis of mesenteric vein	Condition	SNOMED
762049	Acute thrombosis of right subclavian vein	Condition	SNOMED
36712892	Acute thrombosis of splenic vein	Condition	SNOMED
4179911	Axillary vein thrombosis	Condition	SNOMED
44782762	Acute thrombosis of subclavian vein	Condition	SNOMED
37109253	Bilateral acute deep vein thrombosis of femoral veins	Condition	SNOMED
40478951	Bilateral deep vein thrombosis of lower extremities	Condition	SNOMED
4042396	Deep thrombophlebitis	Condition	SNOMED
77310	Deep vein phlebitis and thrombophlebitis of the leg	Condition	SNOMED
4133004	Deep venous thrombosis	Condition	SNOMED
45773536	Deep venous thrombosis of femoropopliteal vein	Condition	SNOMED
763942	Deep venous thrombosis of left lower extremity	Condition	SNOMED
761980	Deep venous thrombosis of left upper extremity	Condition	SNOMED
443537	Deep venous thrombosis of lower extremity	Condition	SNOMED
4133975	Deep venous thrombosis of pelvic vein	Condition	SNOMED
40480555	Deep venous thrombosis of peroneal vein	Condition	SNOMED
4322565	Deep venous thrombosis of profunda femoris vein	Condition	SNOMED
763941	Deep venous thrombosis of right lower extremity	Condition	SNOMED
761928	Deep venous thrombosis of right upper extremity	Condition	SNOMED
4207899	Deep venous thrombosis of tibial vein	Condition	SNOMED
4028057	Deep venous thrombosis of upper extremity	Condition	SNOMED
435565	Embolism and thrombosis of the vena cava	Condition	SNOMED
40481089	Embolism from thrombosis of vein of lower extremity	Condition	SNOMED
4119760	Iliofemoral deep vein thrombosis	Condition	SNOMED
4124856	Inferior mesenteric vein thrombosis	Condition	SNOMED
4096099	Phlebitis of deep veins of lower extremity	Condition	SNOMED
4281689	Phlegmasia alba dolens	Condition	SNOMED

Concept ID	Concept name	domain	vocabulary
4284538	Phlegmasia cerulea dolens	Condition	SNOMED
46285905	Provoked deep vein thrombosis	Condition	SNOMED
46271900	Recurrent deep vein thrombosis	Condition	SNOMED
4033521	Splenic vein thrombosis	Condition	SNOMED
4055089	Superior mesenteric vein thrombosis	Condition	SNOMED
4230403	Thrombophlebitis of axillary vein	Condition	SNOMED
4069561	Thrombophlebitis of deep femoral vein	Condition	SNOMED
761831	Thrombophlebitis of deep vein of bilateral lower limbs	Condition	SNOMED
761830	Thrombophlebitis of deep vein of left lower limb	Condition	SNOMED
761808	Thrombophlebitis of deep vein of left upper limb	Condition	SNOMED
761832	Thrombophlebitis of deep vein of right lower limb	Condition	SNOMED
761809	Thrombophlebitis of deep vein of right upper limb	Condition	SNOMED
4221821	Thrombophlebitis of deep veins of lower extremity	Condition	SNOMED
440750	Thrombophlebitis of deep veins of upper extremities	Condition	SNOMED
4176614	Thrombophlebitis of iliac vein	Condition	SNOMED
761821	Thrombophlebitis of left deep femoral vein	Condition	SNOMED
761819	Thrombophlebitis of left femoral vein	Condition	SNOMED
761820	Thrombophlebitis of right deep femoral vein	Condition	SNOMED
761818	Thrombophlebitis of right femoral vein	Condition	SNOMED
4110339	Thrombophlebitis of the anterior tibial vein	Condition	SNOMED
4111868	Thrombophlebitis of the common iliac vein	Condition	SNOMED
4110343	Thrombophlebitis of the external iliac vein	Condition	SNOMED
439314	Thrombophlebitis of the femoral vein	Condition	SNOMED
4109877	Thrombophlebitis of the internal iliac vein	Condition	SNOMED
4112171	Thrombophlebitis of the popliteal vein	Condition	SNOMED
4112172	Thrombophlebitis of the posterior tibial vein	Condition	SNOMED
4250765	Thrombophlebitis of tibial vein	Condition	SNOMED
42538533	Thrombosis of iliac vein	Condition	SNOMED
44811347	Thrombosis of internal jugular vein	Condition	SNOMED
765049	Thrombosis of left peroneal vein	Condition	SNOMED
4317289	Thrombosis of mesenteric vein	Condition	SNOMED
4203836	Thrombosis of subclavian vein	Condition	SNOMED
4175649	Thrombosis of the popliteal vein	Condition	SNOMED
4149782	Thrombosis of vein of lower limb	Condition	SNOMED
46285904	Unprovoked deep vein thrombosis	Condition	SNOMED
199837	Portal vein thrombosis	Condition	SNOMED
193512	Embolism and thrombosis of the renal vein	Condition	SNOMED

Concept ID	Concept name	domain	vocabulary
4111861	Phlebitis of popliteal vein	Condition	SNOMED
440738	Phlebitis of the femoral vein	Condition	SNOMED
Pulmonary embolism			
43530605	Pulmonary embolism with pulmonary infarction	Condition	SNOMED
45768439	Acute pulmonary embolism	Condition	SNOMED
45768888	Acute pulmonary thromboembolism	Condition	SNOMED
440417	Pulmonary embolism	Condition	SNOMED
4120091	Acute massive pulmonary embolism	Condition	SNOMED
4121618	Pulmonary thromboembolism	Condition	SNOMED
37016922	Pulmonary embolism on long-term anticoagulation therapy	Condition	SNOMED
4253796	Pulmonary microemboli	Condition	SNOMED
35615055	Saddle embolus of pulmonary artery with acute cor pulmonale	Condition	SNOMED
40761384	Pulmonary embolism requiring hospitalization	Observation	LOINC
36713113	Saddle embolus of pulmonary artery	Condition	SNOMED
4119607	Subacute massive pulmonary embolism	Condition	SNOMED
762808	Infarction of lung due to embolus	Condition	SNOMED

ANNEX V. Feasibility counts

Table S2. Descendant Patient Counts of antithrombotics (ATC B01) in participating data sources.

Ingredient name	CPRD GOLD	DK-DHR	FinOMOP-THL	IPCI	SIDIAP
Ingredients included for ingredient-level analyses					
Acenocoumarol	2000			56200	255000
Alteplase	100	25700	300		
Apixaban	131800	193300	269200	37700	100400
Aspirin	1769100	1515600	643900	159800	868000
Bemiparin	100	1700			186400
carbasalate				57300	
Cilostazol	3700				20900
Clopidogrel	443400	443900	220600	87100	219500
dabigatran	8700	60900	47800	11000	32800
Dalteparin	63100	80400	25900	19700	26300
Dipyridamole	100600	159300	63400	17400	3200
Edoxaban	30000	17600	31800	9700	38700
Enoxaparin	87200	9900	581700	3200	919400
Fondaparinux	500	200	1500	2600	6000
Heparin	10800	78600	700	1100	10200
Nadroparin				58600	6000
phenprocoumon		31300	100	11700	
Prasugrel	9300	18100	3400	3200	10500
Rivaroxaban	84300	237700	224100	30200	62800
Streptokinase	1800	8400			280000
Sulodexide					59100
Ticagrelor	26800	54600	51500	14900	23500
Tinzaparin	19600	27600	211000	8200	41600
Triflusal					28300
warfarin	356200	295600	238800	100	25200
Ingredients not included for ingredient-level analyses (drug class analyses only)					
abciximab			100		
aloxiprin	100				
anistreplase	100				
antithrombin III		100			
argatroban					
bivalirudin			100		
cangrelor			100		
caplacizumab					

Ingredient name	CPRD GOLD	DK-DHR	FinOMOP-THL	IPCI	SIDIAP
danaparoid			100		
defibrotide		100			
desirudin					
dicumarol					
drotrecogin alfa					
epoprostenol	100				
eptifibatide			100		
fluidione					
iloprost	100	100	200	100	
indobufen					
lepirudin					
parnaparin					
phenindione	800	100			
plasmin					
protein C					
reteplase					
reviparin					
selexipag			100	100	
tenecteplase	100	1200			
ticlopidine	100				1800
tirofiban			100		
treprostinil		100	100		
urokinase	100				
vorapaxar					
Ximelagatran		100			

ANNEX VI. Mock tables and figures

Table S3. Mock Table: Characteristics of baseline comorbidities and initial treatment of individuals with VTE, in men and women.

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
Women							
Number records	–	N					
Number of individuals	–	N					
Age	–	Median (IQR)					
Age group	18 – 29	N (%)					
	30 – 49	N (%)					
	50 – 64	N (%)					
	65+	N (%)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					
	Apixaban	N (%)					
	...	N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [-365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					
	Extreme under/overweight	N (%)					

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
Comedication [-365, 0]	Use of hormonal contraceptives	N (%)					
Time to treatment initiation	–	Median (IQR)					
Men							
Number records	–	N					
Number of individuals	–	N					
Age	–	Median (IQR)					
Age group	18 – 29	N (%)					
	30 – 49	N (%)					
	50 – 64	N (%)					
	65+	N (%)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					
	Apixaban	N (%)					
	...	N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [-365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
	Extreme under/overweight	N (%)					
Comedication [-365, 0]	Use of hormonal contraceptives	N (%)					
Time to treatment initiation	–	Median (IQR)					

CPRD GOLD = Clinical Practice Research Datalink GOLD; DK-DHR = Danish Data Health Registries; FinOMOP-THL = Finnish Care Register for Health Care; IPCI = Integrated Primary Care Information; IQR = interquartile range; SIDIAP = The Information System for Research in Primary Care; VTE = Venous thromboembolism.

Table S4. Mock Table: Characteristics of baseline comorbidities and initial treatment of individuals with VTE, in men and women, stratified by age group.

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
Women; 18 to 29 years							
Number records	–	N					
Number of individuals	–	N					
Age	–	Median (IQR)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					
	Apixaban						
	...						
	Combination 1						
	Combination 2						
	...						
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [-365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
	Extreme under/overweight	N (%)					
Comedication [-365, 0]	Use of hormonal contraceptives	N (%)					
Time to treatment initiation	–	Median (IQR)					
Men; 18 to 29 years							
Number records	–	N					
Number of individuals	–	N					
Age	–	Median (IQR)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					
	Apixaban	N (%)					
	...						
	Combination 1						
	Combination 2						
	...						
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [-365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					
	Extreme under/overweight	N (%)					
Time to treatment initiation	–	Median (IQR)					
Women; 30 to 49 years							
Number records	–	N					

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
Number of individuals	–	N					
Age	–	Median (I QR)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					
	Apixaban	N (%)					
	...	N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [- 365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					
	Extreme under/overweight	N (%)					
Comedication [- 365, 0]	Use of hormonal contraceptives	N (%)					
Time to treatment initiation	–	Median (I QR)					
Men; 30 to 49 years							
Number records	–	N					
Number of individuals	–	N					
Age	–	Median (I QR)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
	Apixaban	N (%)					
	...	N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [-365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					
	Extreme under/overweight	N (%)					
Time to treatment initiation	–	Median (IQR)					
Women; 50 to 64 years							
Number records	–	N					
Number of individuals	–	N					
Age	–	Median (IQR)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					
	Apixaban						
	...						
	Combination 1						
	Combination 2						
	...						
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [-365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					
	Extreme under/overweight	N (%)					
Comedication [-365, 0]	Use of hormonal contraceptives	N (%)					
Time to treatment initiation	–	Median (IQR)					
Men; 50 to 64 years							
Number records	–	N					
Number of individuals	–	N					
Age	–	Median (IQR)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					
	Apixaban	N (%)					
	...						
	Combination 1						
	Combination 2						
	...						
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [-365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
	Extreme under/overweight	N (%)					
Time to treatment initiation	–	Median (IQR)					
Women; 65+ years							
Number records	–	N					
Number of individuals	–	N					
Age	–	Median (IQR)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					
	Apixaban	N (%)					
	...	N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [-365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					
	Extreme under/overweight	N (%)					
Comedication [-365, 0]	Use of hormonal contraceptives	N (%)					
Time to treatment initiation	–	Median (IQR)					
Men; 65+ years							
Number records	–	N					

Variable name	Variable level	Estimate name	Data source				
			DK-DHR	FinOMOP-THL	SIDIAP	IPCI	CPRD GOLD
Number of individuals	–	N					
Age	–	Median (IQR)					
Initial treatment (ingredient-level)	Untreated	N (%)					
	Acenocoumarol	N (%)					
	Alteplase	N (%)					
	Apixaban	N (%)					
	...	N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
Initial treatment (class-level)	Untreated	N (%)					
	B01AA	N (%)					
	B01AB	N (%)					
		N (%)					
	Combination 1	N (%)					
	Combination 2	N (%)					
	...	N (%)					
Comorbidities [-365, 0]	Chronic kidney disease	N (%)					
	Active cancer	N (%)					
	Extreme under/overweight	N (%)					
Time to treatment initiation	–	Median (IQR)					

CPRD GOLD = Clinical Practice Research Datalink GOLD; DK-DHR = Danish Data Health Registries; FinOMOP-THL = Finnish Care Register for Health Care; IPCI = Integrated Primary Care Information; IQR = interquartile range; SIDIAP = The Information System for Research in Primary Care; VTE = Venous thromboembolism.

Table S5. Mock Table: Drug utilisation of initial treatment for VTE, in men and women.

Ingredient	Estimate name	Data source									
		DK-DHR		FinOMOP-THL		SIDIAP		IPCI		CPRD GOLD	
		Variable name									
		Initial exposure duration	Initial daily dose (milligram)	Initial exposure duration	Initial daily dose (milligram)	Initial exposure duration	Initial daily dose (milligram)	Initial exposure duration	Initial daily dose (milligram)	Initial exposure duration	Initial daily dose (milligram)
Women											
ingredient 1	Median (IQR)										
ingredient 2	Median (IQR)										
ingredient 3	Median (IQR)										
ingredient 4	Median (IQR)										
...											
Men											
ingredient 1	Median (IQR)										
ingredient 2	Median (IQR)										
ingredient 3	Median (IQR)										

Ingredient	Estimate name	Data source									
		DK-DHR		FinOMOP-THL		SIDIAP		IPCI		CPRD GOLD	
		Variable name									
		Initial exposure duration	Initial daily dose (milligram)	Initial exposure duration	Initial daily dose (milligram)	Initial exposure duration	Initial daily dose (milligram)	Initial exposure duration	Initial daily dose (milligram)	Initial exposure duration	Initial daily dose (milligram)
ingredient 4	Median (IQR)										
...											

CPRD GOLD = Clinical Practice Research Datalink GOLD; DK-DHR = Danish Data Health Registries; FinOMOP-THL = Finnish Care Register for Health Care; IPCI = Integrated Primary Care Information; IQR = interquartile range; SIDIAP = The Information System for Research in Primary Care; VTE = Venous thromboembolism.

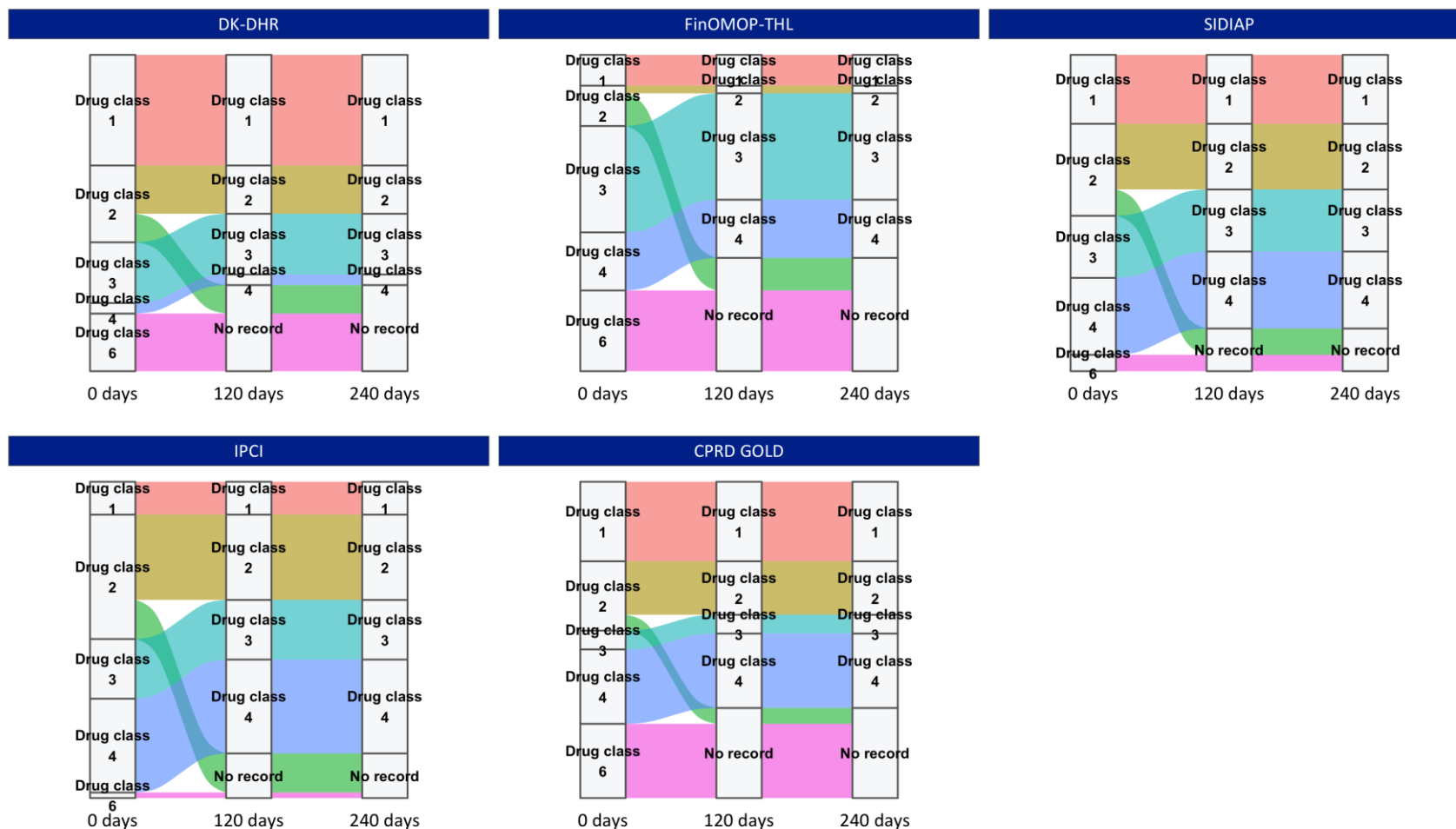


Figure S1. Mock Figure: Sankey plot for treatment switching patterns at drug-class level in men, per data source.

CPRD GOLD = Clinical Practice Research Datalink GOLD; DK-DHR = Danish Data Health Registries; FinOMOP-THL = Finnish Care Register for Health Care; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care.

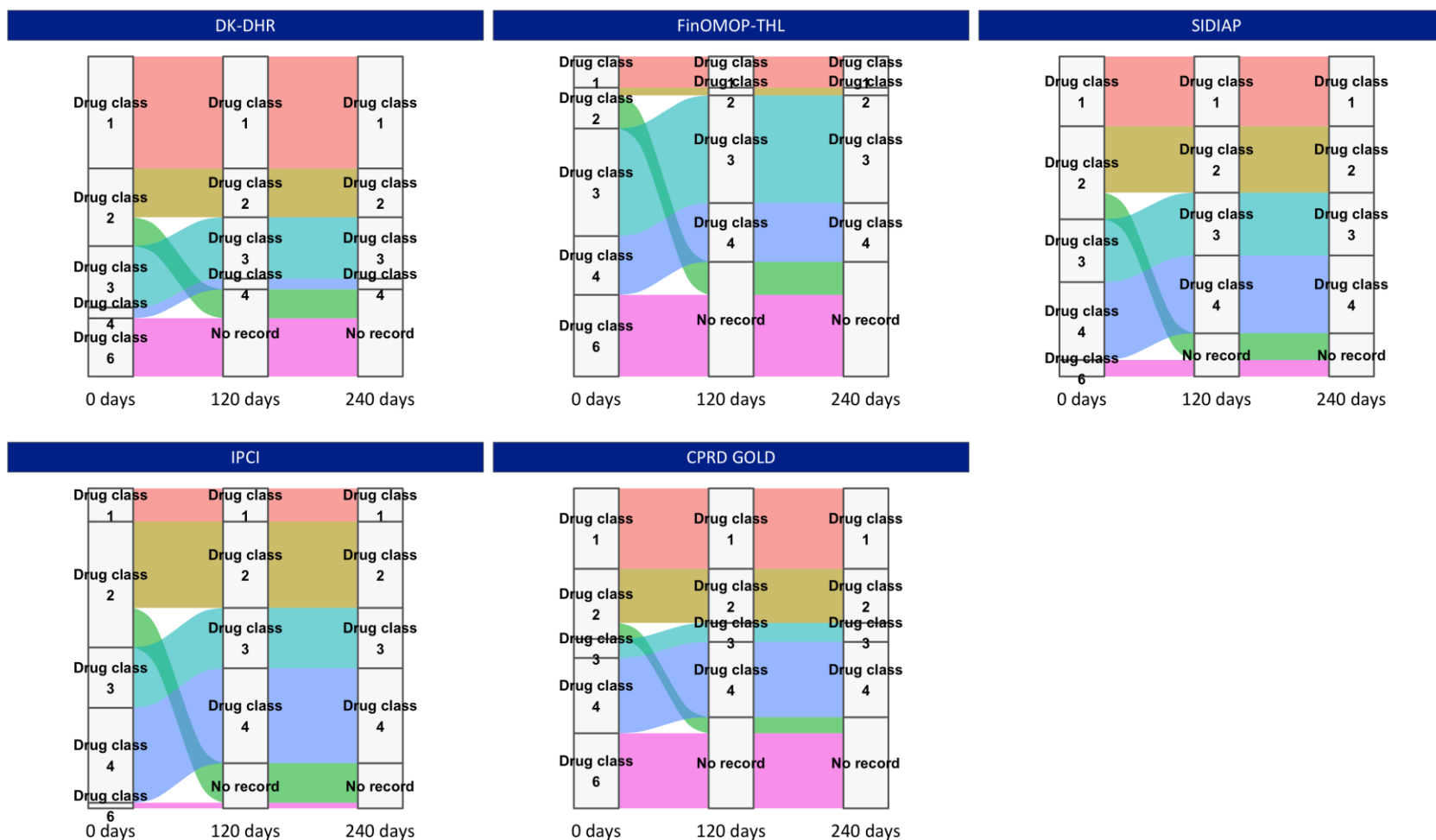


Figure S2. Mock Figure: Sankey plot for treatment switching patterns at drug-class level in women, per data source.

CPRD GOLD = Clinical Practice Research Datalink GOLD; DK-DHR = Danish Data Health Registries; FinOMOP-THL = Finnish Care Register for Health Care; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care.

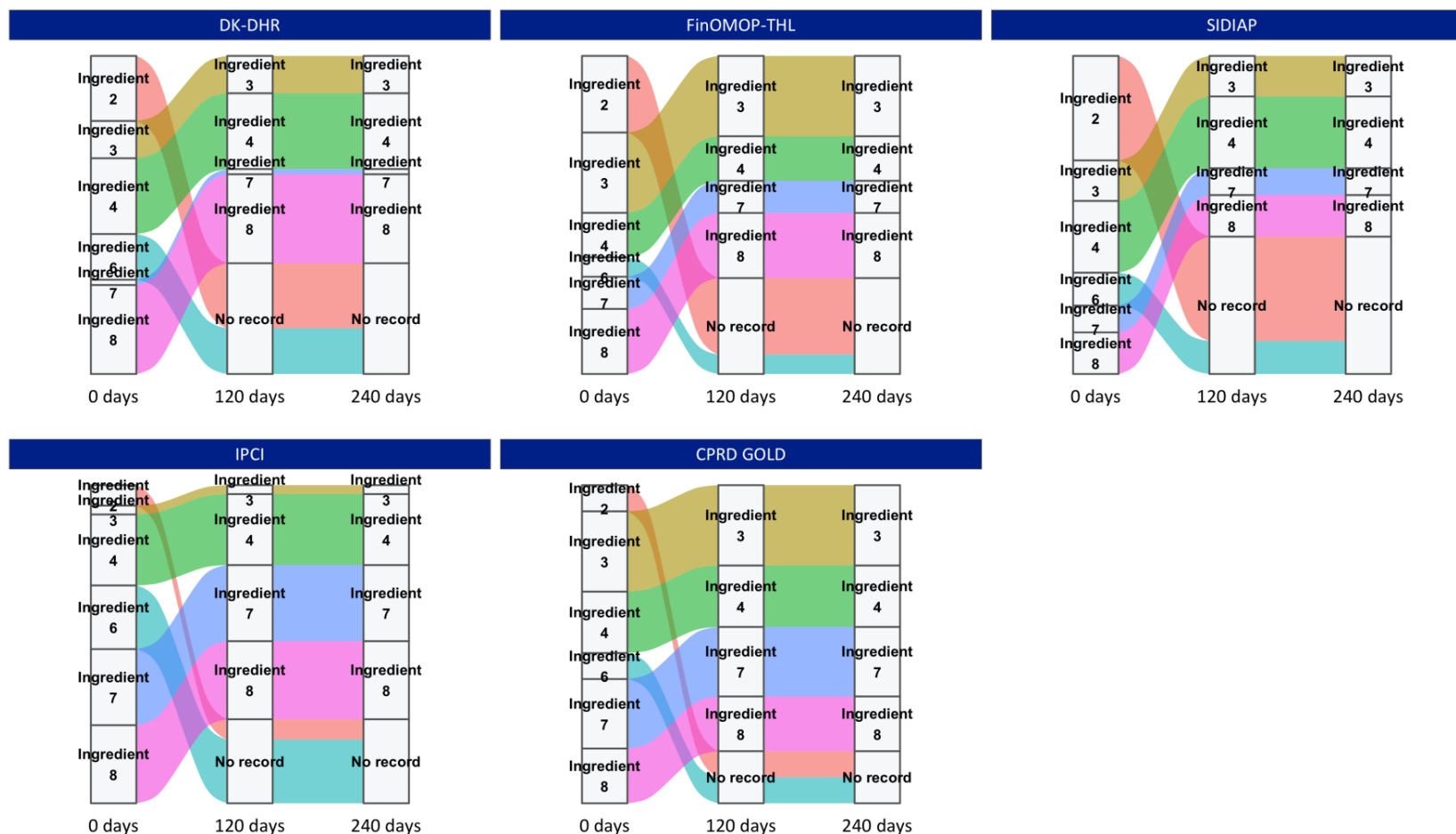


Figure S3. Mock Figure: Sankey plot for treatment switching patterns at ingredient level in men, per data source.

CPRD GOLD = Clinical Practice Research Datalink GOLD; DK-DHR = Danish Data Health Registries; FinOMOP-THL = Finnish Care Register for Health Care; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care.

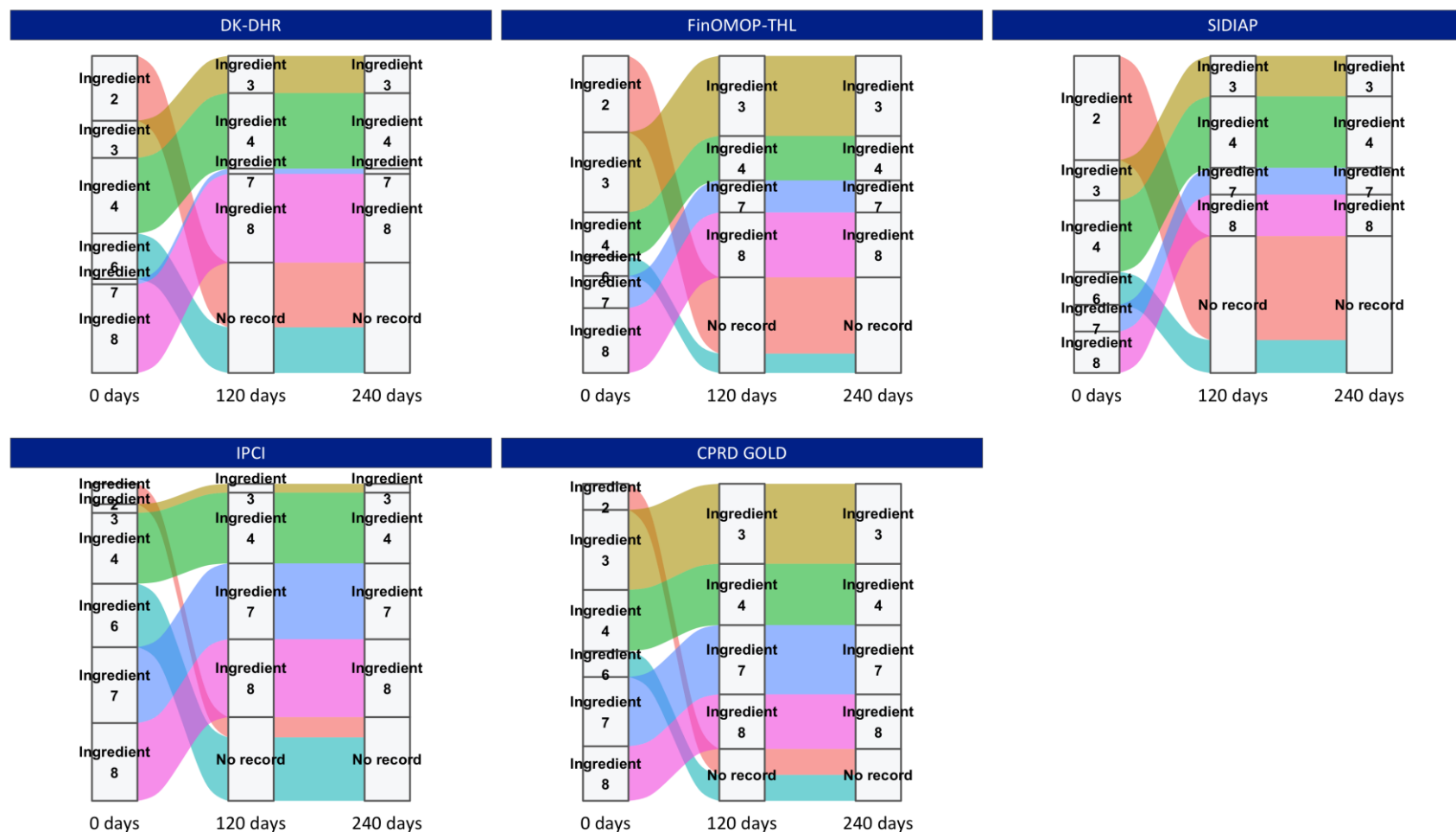


Figure S4. Mock Figure: Sankey plot for treatment switching patterns at ingredient level in women, per data source.

CPRD GOLD = Clinical Practice Research Datalink GOLD; DK-DHR = Danish Data Health Registries; FinOMOP-THL = Finnish Care Register for Health Care; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care.

ANNEX VII. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Sex differences in treatment patterns for venous thromboembolic events in European countries

EU PAS Register® number: EUPAS1000001079

Study reference number (if applicable): P5-C1-006

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	☒	☐	☐	8.2
1.1.2 End of data collection ²	☒	☐	☐	
1.1.3 Progress report(s)	☐	☐	☒	
1.1.4 Interim report(s)	☐	☐	☒	
1.1.5 Registration in the EU PAS Register®	☒	☐	☐	5
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)	☒	☐	☐	7
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)	☒	☐	☐	8.3
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no <i>a priori</i> hypothesis?	☐	☐	☒	

Comments:

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

² Date from which the analytical dataset is completely available.

Section 3: Study design		Yes	No	N/A	Section Number
3.1	Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	8.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	8.4
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☐	☐	☒	
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:				8.3
4.2.1	Study time period	☒	☐	☐	
4.2.2	Age and sex	☒	☐	☐	
4.2.3	Country of origin	☐	☐	☒	
4.2.4	Disease/indication	☒	☐	☐	
4.2.5	Duration of follow-up	☒	☐	☐	
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	8.3

Comments:

Section 5: Exposure definition and measurement		Yes	No	N/A	Section Number
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	8.6
5.2	Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☐	☒	
5.3	Is exposure categorised according to time windows?	☒	☐	☐	8.6

<u>Section 5: Exposure definition and measurement</u>	Yes	No	N/A	Section Number
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☐	☐	☒	8.1
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:

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

Comments:

<u>Section 7: Bias</u>	Yes	No	N/A	Section Number
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☐	☐	☒	
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☐	☐	☒	
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☐	☐	☒	

Comments:

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

Comments:

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

Comments:

<u>Section 10: Analysis plan</u>	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?	☐	☐	☒	

<u>Section 10: Analysis plan</u>	Yes	No	N/A	Section Number
10.3 Are descriptive analyses included?	☒	☐	☐	8.8
10.4 Are stratified analyses included?	☒	☐	☐	8.8
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	
10.7 Does the plan describe methods for handling missing data?	☐	☐	☒	
10.8 Are relevant sensitivity analyses described?	☐	☐	☒	

Comments:

<u>Section 11: Data management and quality control</u>	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:

<u>Section 12: Limitations</u>	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? 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).	☒ ☒ ☐	☐ ☐ ☐	☐ ☐ ☒	9
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 V, Separate Feasibility assessment documented

Comments:

<u>Section 13: Ethical/data protection issues</u>	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 III

Comments:

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

Comments:

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

Comments:

Name of the main author of the protocol: Annika Jodicke

Date: 30/June/2026

Signature: A. Jodicke

ANNEX VIII. Glossary

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

Aggregated Data

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

Benefit-Risk Assessment

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

Common Data Model (CDM)

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

Complex Studies (C3)

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

Coordination Centre (CC)

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

Data Access

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

Data Quality Framework

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

Data Source

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

DARWIN EU®

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

EMA (European Medicines Agency)

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

Evidence Generation

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

Federated Network

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

GDPR (General Data Protection Regulation)

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

Health Technology Assessment (HTA)

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

Metadata

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

Off-the-Shelf Studies (OTS)

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

OHDSI (Observational Health Data Sciences and Informatics)

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

Patient-Level Data

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

OMOP (Observational Medical Outcomes Partnership)

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

Real-World Data (RWD)

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

Real-World Evidence (RWE)

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

Regulatory Decision-Making

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

Routine Repeated Studies (RR)

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

Study Protocol

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

Very Complex Studies (C4)

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