Study Protocol

P4-C2-017

DARWIN EU® - Time to onset of thromboembolic events in adults with selected types of cancer

10/12/2025

Version 3.0

Authors: Melissa Leung, Cesar Barboza, Ionna Nika, Anton Barchuk, Talita Duarte-Salles

Public

CONTENTS

Study title	DARWIN EU® - Time to onset of thromboembolic events in adults with selected types of cancer
Protocol version	V3.0
Date	10/12/2025
EUPAS number	EUPAS1000000814
Active substance	Not applicable
Medicinal product	Not applicable
Objectives	The aim of this study is to estimate time to onset of venous thromboembolic events in adults with each type of selected cancer. The specific objectives of the study are: 1. To estimate the cumulative incidence of thromboembolism within 5 years after the first cancer diagnosis in adults with each type of selected cancer, overall and stratified by age group, sex, and study subperiod. 2. To estimate the median time from the first cancer diagnosis to onset of thromboembolic events in individuals with thromboembolic events with each type of selected cancer, overall and stratified by age group, sex, and study subperiod.
Countries of study	Belgium, Denmark, Estonia, Finland, Germany, The Netherlands, Spain, United Kingdom
Authors	Melissa Leung (m.leung@darwin-eu.org) Cesar Barboza (c.barboza@darwin-eu.org) Ioanna Nika (i.nika@darwin-eu.org) Anton Barchuk (a.barchuk@darwin-eu.org) Talita Duarte-Salles (t.duarte@darwin-eu.org)

This is a routinely repeated study from P3-C3-005 with EUPAS1000000440.

LIST OF ABBREVIATIONS

Acronyms/terms	Description
AJCC/UICC	American Joint Committee on Cancer and the International Union Against Cancer
ATC	Anatomical Therapeutic Chemical
CDM	Common Data Model
CI	Confidence interval
CPRD	Clinical Practice Research Datalink
DA	Disease Analyser
DARWIN EU®	Data Analysis and Real World Interrogation Network
DK-DHR	Danish Data Health Registries
DTZ	Data Transfer Zone
DOI	Declaration Of Interests
DQD	Data Quality Dashboard
DRE	Digital Research Environment
DVT	Deep Venous Thrombosis
DIC	Disseminated Intravascular Coagulation
EHR	Electronic Health Record
EMA	European Medicines Agency
EBB	Estonian Biobank
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
k	Thousands
ICD-O-3	International Classification of Diseases for Oncology, 3rd Edition
ICD-10	International Classification of Diseases, 10th revision
ICPC-1	International Classification of Primary Care
IP	Inpatient
IPCI	Integrated Primary Care Information Project
IRB	Institutional Review Board
LPD	Longitudinal Patient Database
M	Millions
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
PE	Pulmonary Embolism

Acronyms/terms	Description
PVT	Pelvic Venous Thrombosis
RVT	Retinal vein thrombosis
SIDIAP	The Information System for Research on Primary Care
SNOMED	Systematized Nomenclature of Medicine
SVT	Splanchnic Vein Thrombosis
UKBB	UK Biobank
VTE	Venous Thromboembolism

1. TITLE

DARWIN EU® - Time to onset of thromboembolic events in adults with selected types of cancer

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigators	Melissa Leung Anton Barchuk	Erasmus MC
Co-principal Investigator	Talita Duarte-Salles	Erasmus MC
Data Scientists	Cesar Barboza Ioanna Nika Maarten van Kessel Ger Inberg Adam Black Ross Williams	Erasmus MC
Clinical Domain Expert	Anton Barchuk	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	**Names**	**Data Partner Organisation***
IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium) IQVIA Disease Analyzer Germany (IQVIA DA Germany)	Gargi Jadhav Isabella Kacmarczy Akram Mendez Hanne van Ballegooijen Dina Vojinovic	IQVIA
Danish Data Health Registries (DK-DHR)	Elvira Bräuner Susanne Bruun	Danish Medicines Agency
Estonian Biobank (EBB)	Marek Oja Raivo Kolde Ami Sild	Estonian Biobank, Estonia
Finnish Care Register for Health Care (FinOMOP-THL)	Anna Hammais Gustav Klingstedt	Finnish Care Register for Health Care, Finland
Integrated Primary Care Information (IPCI)	Katia Verhamme	Integrated Primary Care Information, Netherlands
The Information System for the Development of Research in Primary Care (SIDIAP)	Anna Palomar-Cros Irene López-Sánchez Agustina Giuliodori	IDIAPJGol
Clinical Practice Research Datalink GOLD (CPRD GOLD) UK BioBank (UKBB)	Antonella Delmestri	University of Oxford

*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® - Time to onset of thromboembolic events in adults with selected types of cancer

Rationale and background

Thromboembolic events are a common complication for individuals with cancer, with risk varying according to the cancer type, suggesting that mechanisms which play a role in the occurrence of these events may be specific to the cancer type itself or its treatment mechanisms. Haematological malignancies and lung, pancreas, stomach, bowel, and brain cancers are generally associated with a high risk of thrombosis, whilst prostate and breast cancers are associated with low risk of thrombosis.

When a safety signal of a thromboembolic event appears in cancer populations, it can be challenging to assess a potential association with the oncologic treatment without reliable information on the background risk. This study is intended to address this knowledge gap by generating evidence on the time to onset of different venous thromboembolic events among adults with selected cancer types.

Objectives

The aim of this study is to estimate time to onset of venous thromboembolic events in adults with each type of selected cancer.

The specific objectives of the study are:

1. To estimate the cumulative incidence of thromboembolism within 5 years after the first cancer diagnosis in adults with each type of selected cancer, overall and stratified by age group, sex, and study subperiod.
2. To estimate the median time from the first cancer diagnosis to onset of thromboembolic events in individuals with thromboembolic events with each type of selected cancer, overall and stratified by age group, sex, and study subperiod.

Methods

Study design

Population-based cohort study. The index date is the date of the first cancer diagnosis. Individuals are followed up until the earliest of occurrence of a thromboembolic event, loss to follow-up, end of data availability, end of the study period, or death.

Population

The study population will include all individuals aged 18 years and above with a primary diagnosis of one of the selected cancers (bone, brain, breast, colorectal, corpus uteri, kidney, leukaemia and lymphoma, liver, lung, melanoma, oesophageal, ovary, pancreas, prostate, stomach) during the inclusion period (from 01/01/2016 to 31/12/2022). Only individuals with an incident cancer diagnosis (excluding non-melanoma skin cancer), defined as a first cancer diagnosis after ≥365 days cancer-free history, will be included. Cancer cases and thromboembolic events will be identified based on appropriate computable phenotyping algorithms. Conditions in the OMOP CDM use the Systematised Nomenclature of Medicine (SNOMED) as the standard vocabulary for diagnosis codes. The International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3) will also be considered for cancer diagnoses. Other eligibility criteria will include at least 365 days of database history prior to index date and at least 365 days between index date and end of data availability in the data source.

Variables

Outcome:

The specific thromboembolic outcomes of interest are: deep vein thrombosis (DVT), pulmonary embolism (PE), venous thromboembolism (VTE, composite of DVT and PE), pelvic venous thrombosis (PVT), splanchnic vein thrombosis (SVT, including hepatic and extra-hepatic vein thrombosis), retinal vein thrombosis (RVT, including retinal central vein thrombosis), and disseminated intravascular coagulation (DIC).

Relevant covariates:

The following covariates will be assessed at index date: age (categorised into age groups: 18–34, 35–44, 45–54, 55–64, 65–74, 75–84, and ≥85 years), sex, and calendar year of the first cancer diagnosis (categorised into study subperiods: 2016–2019 and 2020–2022). These variables will be used to stratify the results.

Data sources

1. Belgium: IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)
2. Denmark: Danish Data Health Registries (DK-DHR)
3. Estonia: Estonian Biobank (EBB)
4. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
5. Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)
6. Netherlands: Integrated Primary Care Information (IPCI)
7. Spain: The Information System for the Development of Research on Primary Care (SIDIAP)
8. United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)
9. United Kingdom: UK BioBank (UKBB)

Study size

No sample size has been calculated, as this is a descriptive study which will not test a specific hypothesis. Based on the results of the study EUPAS1000000440, the expected number of person counts will be the lowest for DIC (during 1-year follow-up: 5 in FinOMOP-THL – 171 in SIDIAP, with 0 counts in CPRD GOLD, EBB, IPCI, IQVIA DA Germany, and IQVIA LPD Belgium) and highest for VTE (during 1-year follow-up: 27 in IQVIA LPD Belgium – 4,597 in FinOMOP-THL).

Statistical analysis

Analyses will be conducted separately for each data source and carried out in a federated manner, allowing analyses to be run locally without sharing individual-level data. The analyses will be conducted overall, as well as by strata of age, sex, and calendar year of the first cancer diagnosis.

Absence of diagnosis codes will be interpreted as the absence of the conditions themselves. A minimum cell count of 5 will be used when reporting results, with any smaller count reported as “<5” and zero counts as “0”.

Objective 1

The cumulative incidence of thromboembolism in adults with each type of selected cancer will be calculated in a competing risk survival analysis and reported at 6-month intervals within 5 years after the first cancer diagnosis, accounting for death as a competing risk.

Objective 2

Among the individuals with cancer who experienced a thromboembolic event, the median time from the first cancer diagnosis to the onset of the thromboembolic event will be calculated.

4. AMENDMENTS AND UPDATES

None

5. MILESTONES

Study milestones and deliverables	Planned dates
Final Study Protocol	10 December 2025
Creation of Analytical code	December 2025
Execution of Analytical Code on the data	December 2025
Draft Study Report	January 2026
Final Study Report	February 2026

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

6. RATIONALE AND BACKGROUND

Thromboembolic events are a common complication for individuals with cancer, with risk varying according to the cancer type, suggesting that mechanisms that play a role in the occurrence of these events may be specific to the cancer type itself or its treatment.[1,2] Haematological malignancies and lung, pancreas, stomach, bowel, and brain cancers are generally associated with a high risk of thrombosis, whilst prostate and breast cancers are associated with low risk of thrombosis.[3]

When a safety signal of a thromboembolic event appears in cancer populations, it can be challenging to assess a potential association with the oncologic treatment without reliable information on the background risk. This study is intended to address this knowledge gap by generating evidence on the time to onset of different venous thromboembolic events among adults with selected cancer types.

This study is a routine repeated study of a previous DARWIN EU® study (EUPAS1000000440) that focused on estimating the incidence rates of venous thromboembolic events (VTE) in adults newly diagnosed with any of the following selected cancers (bone, brain, breast, colorectal, corpus uteri, kidney, leukaemia and lymphoma, liver, lung, melanoma, oesophageal, ovary, pancreas, prostate, stomach) during the period 2016–2022 and describing the individuals' characteristics at the time of cancer diagnosis. This study is now being repeated with the same study population and outcome, but with different objectives: to estimate the cumulative incidence of thromboembolism and the time to onset of thromboembolic events in adults with these selected types of cancer.

7. RESEARCH QUESTION AND OBJECTIVES

Objectives

The aim of this study is to estimate time to onset of venous thromboembolic events in adults with each type of selected cancer.

The specific objectives of the study are:

1. To estimate the cumulative incidence of thromboembolism within 5 years after the first cancer diagnosis in adults with each type of selected cancer, overall and stratified by age group, sex, and study subperiod.
2. To estimate the median time from the first cancer diagnosis to onset of thromboembolic events in individuals with thromboembolic events with each type of selected cancer, overall and stratified by age group, sex, and study subperiod.

8. RESEARCH METHODS

8.1. Study design

A cohort study will be conducted.

8.2. Follow-up

For both objectives, follow-up will start on the date of the first cancer diagnosis (index date) and end on the date of occurrence of the outcome (thromboembolic event), loss to follow-up, end of data availability, 5 years of follow-up, or death, whichever comes first.

8.3. Study population with inclusion and exclusion criteria

Objective 1

Inclusion criteria

- First diagnosis of a selected cancer (index date) between 01/01/2016 and 31/12/2022 (inclusion period)
- Age ≥18 years at cancer diagnosis
- Minimum of 365 days of available history before the cancer diagnosis date
- Cancer diagnosis date ≥365 days prior to end of data availability of the data source

Exclusion criteria

- History of any cancer diagnosis ever before the selected cancer diagnosis date

Objective 2

Inclusion criteria

- The subset of the cohort for objective 1 who experienced the outcome (i.e., thromboembolic event) during follow-up.

The study design to address objective 1, including assessment windows, is visualised in **Figure 1**. For objective 2, we will include the subset of the cohort of adults with cancer that experienced a thromboembolic event during follow-up.

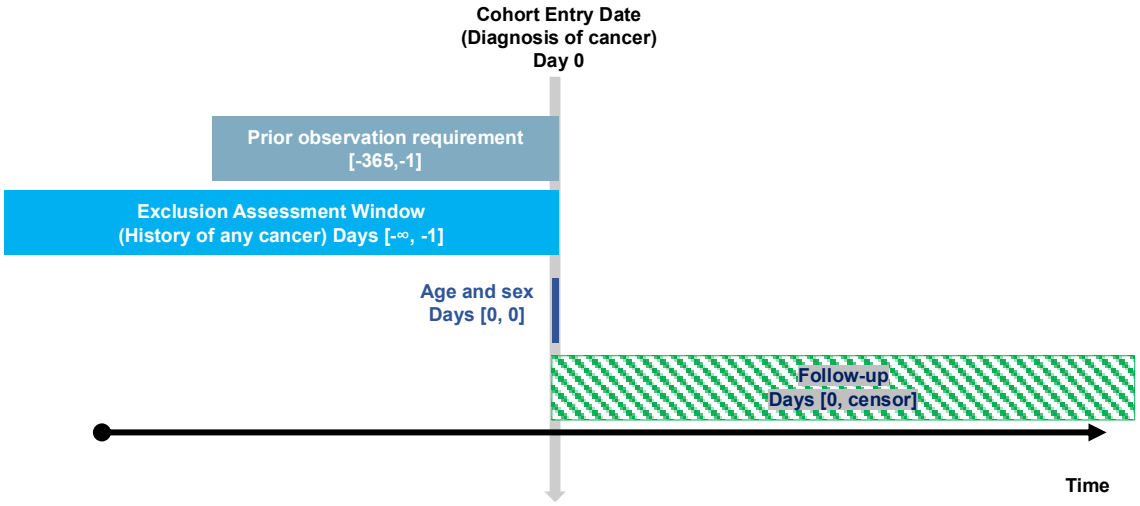

Figure 1. Graphical depiction of the study design.

The censor date will be the earliest of occurrence of the outcome, loss to follow-up, end of data availability, or death.

8.4. Study setting and data sources

This study will be conducted using data collected in routine care from different health care settings from 9 data sources part of the DARWIN EU® network and distributed in 8 different countries across Europe (**Table 1**). All data were a priori mapped to the OMOP CDM.

Table 1. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to
Belgium	IQVIA LPD Belgium	Primary care	EHRs	189k	2015–2025	All objectives
Denmark	DK-DHR	All settings	EHRs, registries, claims	5.98M	1995–2025	All objectives
Estonia	EBB	Primary care, hospital care (IP and OP)	EHRs, claims, registries, biobank	212k	2004–2025	All objectives
Finland	FinOMOP-THL	Hospital care (IP and OP)	EHRs, registries	5.7M	2011–2025	All objectives
Germany	IQVIA DA Germany	Primary care	EHRs	4.48M	1992–2025	All objectives
The Netherlands	IPCI	Primary care	EHRs	1.33M	2006–2025	All objectives
Spain	SIDIAP	Primary care	EHRs	5.95M	2006–2025	All objectives
United Kingdom of Great Britain and Northern Ireland	CPRD GOLD	Primary care, hospital care (OP)	EHRs	2.83M	1987–2025	All objectives
United Kingdom of Great Britain and Northern Ireland	UKBB	Primary care (up to 2017), hospital care (IP and OP, up to November 2022)	EHRs, registries, biobank	500k	1940–2025	All objectives

Data sources: IQVIA LPD=IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium); DK-DHR=Danish Data Health Registries; EBB=Estonian Biobank; FinOMOP-THL=Finnish Care Register for Health Care; IQVIA DA=IQVIA Disease Analyzer Germany (IQVIA DA Germany); IPCI=Integrated Primary Care Information; SIDIAP=The Information System for the Development of Research in Primary Care; CPRD GOLD=Clinical Practice Research Datalink GOLD; UKBB=UK BioBank
Types of data: EHR=electronic health record, IP=inpatient, OP=outpatient
Number of active subjects: k=thousands, M=millions

Data sources selection

These data sources fulfil the criteria required in terms of data quality, completeness, timeliness, and representativeness for the cohort study while covering different regions of Europe (**Annex II**).

8.5. Study period

The study period is from 01/01/2016 to the most recent data available for each contributing data source.

8.6. Variables

8.6.1. Outcome

All objectives

The specific thromboembolic events of interest are:

- Deep vein thrombosis (DVT)
- Pulmonary embolism (PE)
- Venous thromboembolism (VTE, composite of DVT and PE)
- Pelvic venous thrombosis (PVT)
- Splanchnic vein thrombosis (SVT), including hepatic and extra-hepatic vein thrombosis
- Retinal vein thrombosis (RVT), including retinal central vein thrombosis
- Disseminated intravascular coagulation (DIC)

Each of the specific thromboembolic events will be a binary outcome. Each outcome will be assessed at any diagnosis position in the electronic health record and from any of the care settings in each data source.

The list of concepts was based on SNOMED codes and aligned with previous studies that used OMOP CDM and VTE as an outcome (Burn et al., 2022) and is provided in **Annex IV**.

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

All Objectives

- Sex
    - Female/male
- Age (years) on the date of the first cancer diagnosis, categorised into age groups:
    - 18–34
    - 35–44
    - 45–54
    - 55–64
    - 65–74
    - 75–84
    - ≥85

  Age at index date will be calculated using January 1st of the year of birth as proxy for the actual birthday. Date/month is either not present or cannot be made available for governance reasons. If available, date is often set to first of the month for personal privacy.

- Calendar year of the first cancer diagnosis, categorised into subperiods of the study period:
    - 2016–2019
    - 2020–2022

The status of each covariate will be assessed on the date of the first cancer diagnosis. Each covariate will be used to stratify the results.

8.7. Study size

No sample size has been calculated, as this is a descriptive study which will not test a specific hypothesis. In addition, we will use data from the study population that was included in EUPAS1000000440. Based on the results of the study EUPAS1000000440, the expected number of person counts is the lowest for DIC (during 1-year follow-up: 5 in FinOMOP-THL – 171 in SIDIAP, with 0 counts in CPRD GOLD, EBB, IPCI, IQVIA DA Germany, and IQVIA LPD Belgium) and highest for VTE (during 1-year follow-up: 27 in IQVIA LPD Belgium – 4,597 in FinOMOP-THL).

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. Operational and reporting considerations**), the final package will be released in a version-controlled study repository for execution against all the participating data sources.

8.8.2. Data privacy protection

The data partners will locally execute the analytics against the OMOP CDM in R Studio and review and approve the default aggregated results. They will then be made available to the Principal Investigators and study team in secure online repository of DTZ (Data Transfer Zone) and Digital Research Environment (DRE). 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 database, and will be carried out in a federated manner, allowing analyses to be run locally without sharing individual-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

Objective 1

The cumulative incidence of thromboembolic events at 6-month intervals within 5 years (6, 12, 18, 24, 30, 36, 42, 48, 54, and 60 months) in adults with each type of selected cancer will be calculated based on OMOP CDM mapped data using the R package *CohortSurvival*,[4] developed by DARWIN EU®. The R package *CohortSurvival* is designed to work with data in the OMOP CDM format to extract and summarise survival data. A competing risk analysis will be conducted to account for the risk of death. The 95% confidence intervals (CIs) for probabilities based on the Aalen-Johansen estimator will be calculated using standard errors derived from the infinitesimal jackknife method.[5] Individuals will be censored on the date of loss to follow-up, the end of data availability, or the end of the study period. The *estimateCompetingRiskSurvival*() function will be used in the analysis. The cumulative incidence will be reported with the 95% CI. In addition, the overall survival curve will be plotted.

Objective 2

The median time to onset of venous thromboembolic events will be calculated based on OMOP CDM mapped data using the R package *CohortCharacteristics*,[6] developed by DARWIN EU®. The median time to onset will be reported with the interquartile range.

All objectives

The analyses will be conducted overall as well as by strata of age group, sex, and study subperiod. The absence of diagnosis codes will be interpreted as the absence of the conditions themselves. A minimum cell count of 5 will be used when reporting results, with any smaller count reported as "<5" and zero counts as "0".

Sensitivity analysis

Not applicable.

8.8.4. Output

Output will include a PDF report including an executive summary, and tables and figures. Mock versions of the intended tables and figures are listed below.

Objectives 1 and 2

- Table 1. Attrition of study participants.

Objective 1

- Figure 1. Cumulative incidence of the thromboembolic event after the cancer diagnosis.
    - This figure will be plotted for each combination of cancer type and outcome, faceted by data source. The main report will include 15 such figures: one for each type of cancer. In the meta-analysis results of EUPAS1000000440, of which this is a routinely repeated study, the most common outcome was VTE across all cancer types, except for liver cancer. In liver cancer, the most common outcome was SVT. Therefore, Figure 1 in the main report will be plotted for SVT in liver cancer and for VTE in all other cancer types. The corresponding figures for all outcomes across all cancer types will be available in the Shiny app.
- Table 2. Cumulative incidence of the thromboembolic event at 6-month intervals up to 5 years after the cancer diagnosis.
    - The main report will include the table with cumulative incidence of the thromboembolic event. We will report it for VTE for 14 cancer types and for SVT in liver cancer. The corresponding results of each of the thromboembolic event outcomes will be available in the Shiny app, reported in one table per outcome.

Objective 2

- Table 3. Median time to thromboembolic event after first cancer diagnosis in those with occurrence of a thromboembolic event.

An interactive dashboard (Shiny) will be generated by incorporating all the results (tables and figures) included in the PDF report mentioned above. Specifically, the Shiny will contain:

- Overall results:
    - Objective 1: 15 cancer types * 7 outcomes = 105 figures as presented in **Figure 1**.
    - Objective 1: 7 tables as presented in **Table 2**.
    - Objective 2: 1 table (**Table 3**).

- Stratified results:
  - By age group:
    - Objective 1: 7 age groups * 105 figures = 735 figures
    - Objective 1: 7 age groups * 7 tables = 49 tables
    - Objective 2: 7 age groups * 1 tables = 7 tables
  - By sex:
    - Objective 1: 2 sexes * 105 figures = 210 figures
    - Objective 1: 2 sexes * 7 tables = 14 tables
    - Objective 2: 2 sexes * 1 table = 2 tables
  - By study subperiod:
    - Objective 1: 2 subperiods * 105 figures = 210 figures
    - Objective 1: 2 subperiods * 7 tables = 14 tables
    - Objective 2: 2 subperiods * 1 table = 2 tables

Table 1. Attrition of study participants.

		IQVIA LPD Belgium	DK-DHR	EBB	FinOMOP-THL	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB
Bone cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Brain cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Breast cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Colorectal cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									

		IQVIA LPD Belgium	DK-DHR	EBB	FinOMOP-THL	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB
Corpus uteri cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Oesophageal cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Kidney cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Liver cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Lung cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Lymphoma and Leukaemia	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Melanoma	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Ovarian cancer	Qualifying initial records									

		IQVIA LPD Belgium	DK-DHR	EBB	FinOMOP-THL	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB
	No prior history of cancer									
	Outcome during follow-up									
Pancreatic cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Prostate cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									
Stomach cancer	Qualifying initial records									
	No prior history of cancer									
	Outcome during follow-up									

N=the number of individuals meeting each criterion.
IQVIA LPD=IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium); DK-DHR=Danish Data Health Registries; EBB=Estonian Biobank; FinOMOP-THL=Finnish Care Register for Health Care; IQVIA DA=IQVIA Disease Analyzer Germany (IQVIA DA Germany); IPCI=Integrated Primary Care Information; SIDIAP=The Information System for the Development of Research in Primary Care; CPRD GOLD=Clinical Practice Research Datalink GOLD; UKBB=UK Biobank.

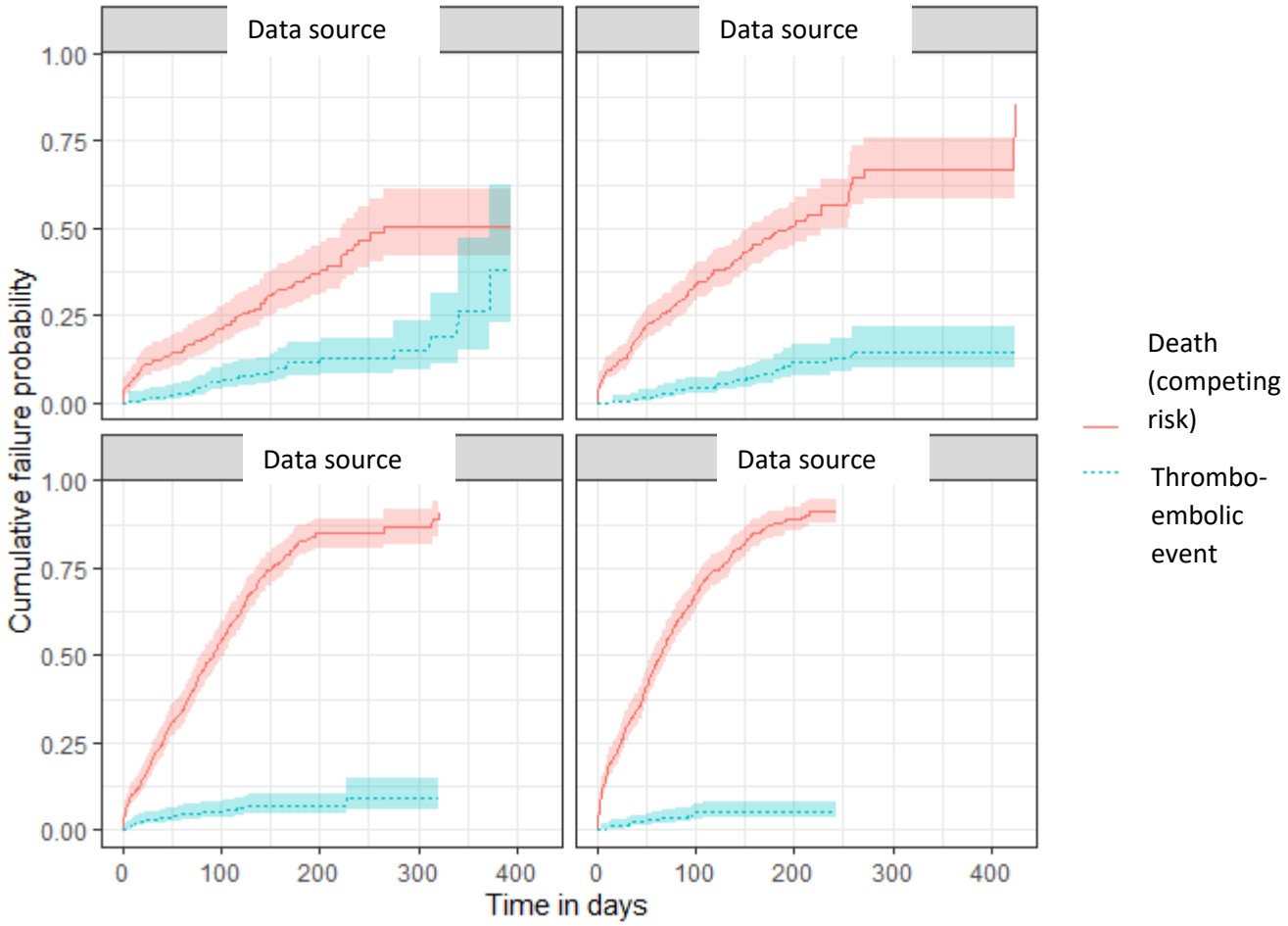

Figure 1. Cumulative incidence of thromboembolic events (SVT in liver cancer and VTE in all other cancer types) after the first cancer diagnosis accounting for a competing risk of death.

Table 2. Cumulative incidence of thromboembolic events (splanchnic vein thrombosis in liver cancer and pulmonary embolism and deep vein thrombosis combined in all other cancer types) at 6-month intervals up to 5 years after the first cancer diagnosis, % (95% confidence interval).

Time since first cancer diagnosis (months)	IQVIA LPD Belgium	DK-DHR	EBB	FinOMOP-THL	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB
Bone cancer (N=[x])	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Brain cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Breast cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									

Time since first cancer diagnosis (months)	IQVIA LPD Belgium	DK-DHR	EBB	FinOMOP-THL	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB
Colorectal cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Corpus uteri cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Oesophageal cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Kidney cancer (N=[x])									

Time since first cancer diagnosis (months)	IQVIA LPD Belgium	DK-DHR	EBB	FinOMOP-THL	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Liver cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Lung cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Lymphoma and Leukaemia (N=[x])									
6									
12									
18									

Time since first cancer diagnosis (months)	IQVIA LPD Belgium	DK-DHR	EBB	FinOMOP-THL	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB
24									
30									
36									
42									
48									
54									
60									
Melanoma (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Ovarian cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Pancreatic cancer (N=[x])									
6									
12									
18									
24									
30									

Time since first cancer diagnosis (months)	IQVIA LPD Belgium	DK-DHR	EBB	FinOMOP-THL	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB
36									
42									
48									
54									
60									
Prostate cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									
Stomach cancer (N=[x])									
6									
12									
18									
24									
30									
36									
42									
48									
54									
60									

IQVIA LPD=IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium); DK-DHR=Danish Data Health Registries; EBB=Estonian Biobank; FinOMOP-THL=Finnish Care Register for Health Care; IQVIA DA=IQVIA Disease Analyzer Germany (IQVIA DA Germany); IPCI=Integrated Primary Care Information; SIDIAP=The Information System for the Development of Research in Primary Care; CPRD GOLD=Clinical Practice Research Datalink GOLD; UKBB=UK BioBank.

Table 3. Median time in days (IQR) to thromboembolic event after first cancer diagnosis.

	IQVIA LPD Belgium	DK-DHR	EBB	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB	FinOMOP-THL
Bone cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Brain cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Breast cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									

	IQVIA LPD Belgium	DK-DHR	EBB	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB	FinOMOP-THL
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Colorectal cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Corpus uteri cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Oesophageal cancer (N=[x])									

	IQVIA LPD Belgium	DK-DHR	EBB	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB	FinOMOP-THL
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Kidney cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Liver cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									

	IQVIA LPD Belgium	DK-DHR	EBB	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB	FinOMOP-THL
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Lung cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Lymphoma and Leukaemia (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Melanoma (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									

	IQVIA LPD Belgium	DK-DHR	EBB	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB	FinOMOP-THL
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Ovarian cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Pancreatic cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									

	IQVIA LPD Belgium	DK-DHR	EBB	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	UKBB	FinOMOP-THL
Disseminated intravascular coagulation (DIC)									
Prostate cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									
Stomach cancer (N=[x])									
Deep vein thrombosis (DVT)									
PE and DVT combined (VTE)									
Pelvic vein thrombosis (PVT)									
Pulmonary embolism (PE)									
Retinal vein thrombosis (RVT)									
Splanchnic extrahepatic vein thrombosis (SVT)									
Disseminated intravascular coagulation (DIC)									

IQVIA LPD=IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium); DK-DHR=Danish Data Health Registries; EBB=Estonian Biobank; FinOMOP-THL=Finnish Care Register for Health Care; IQVIA DA=IQVIA Disease Analyzer Germany (IQVIA DA Germany); IPCI=Integrated Primary Care Information; SIDIAP=The Information System for the Development of Research in Primary Care; CPRD GOLD=Clinical Practice Research Datalink GOLD; UKBB=UK BioBank.

9. STRENGTHS AND LIMITATIONS

The study will be informed by healthcare data collected in routine clinical care for another purpose than the study purpose, therefore data quality issues must be considered. In particular, the identification of individuals with cancer and thromboembolic events may vary across data sources. While relatively few false positives (e.g., individuals wrongly identified as individuals with a cancer diagnosis) are expected, false negatives (e.g., individuals with a cancer diagnosis not identified) may be more likely, especially in primary care data sources that lack patient-level linkage to secondary care data or cancer registry data. We expect misclassification to be minimal in registry data sources and in the primary care data sources that validated information on cancer diagnoses by linking the data to cancer registries, i.e. assuring that cancer cases recorded at primary health sources were later confirmed at specialised facilities (e.g., SIDIAP) [7]. Underestimation of thromboembolic events is also possible, particularly for rare events with complex diagnoses, such as RVT and SVT.

Given the large number and diverse nature of participating data sources, it is important to note that differences in patient representations might be the result of disparate coding practices (ICD-10, ICD-O-3, SNOMED) and specifics of data capture (data collected at the primary healthcare level, data collected at the cancer registry, data collected from the hospital electronic health records and data from the claims). The granularity or detail of concepts representing clinical facts can vary across source terminologies (e.g., ICD-10, Read codes), influencing how information is later transformed into standardised vocabularies.[8] The preliminary code lists created to identify individuals with cancer include codes from standard vocabularies used in cancer registries, such as ICD-O-3 codes. However, most data sources capture information on cancer diagnoses using SNOMED codes, which may not be granular enough to cover all the topology and morphology of cancer.[9] ICD-O-3 codes are only available at DK-DHR, EBB, and UKBB. This limitation restricts our ability to stratify analyses by morphology, which may provide important insights for cancer types encompassing both indolent and aggressive subtypes.

The lack of information on cancer treatment represents a limitation of this study, as it precludes distinguishing between thromboembolic risks related to cancer itself and those potentially associated with cancer treatment.

Additionally, the study period includes the COVID-19 pandemic, which may have influenced both the presentation of cancer cases (with a higher proportion of advanced-stage cancers diagnosed during the pandemic) and the incidence of thromboembolic events. The stratification by calendar period (2016–2019 versus 2020–2022) may provide some insight into this.

10. REFERENCES

1. Wan T, Song J, Zhu D. Cancer-associated venous thromboembolism: a comprehensive review. Thromb J. 2025 Apr 16;23(1):35.

2. Vladić N, Englisch C, Berger JM, Moik F, Berghoff AS, Preusser M, et al. Validation of risk assessment models for venous thromboembolism in patients with cancer receiving systemic therapies. Blood Adv. 2025 June 30;9(13):3340–9.

3. Deng HW, Li J, Zhai YS, Mei WY, Lin XX, Xu Q, et al. Incidence of arterial and venous thromboembolism in cancer patients- insights from more than 5,000,000 patients. J Thromb Thrombolysis. 2025 Mar 1;58(3):370–9.

4. López-Güell K, Burn E, Catala M, Li X, Newby D, Mercade-Besora N (2025). *CohortSurvival: Estimate Survival from Common Data Model Cohorts*. R package version 1.0.3, https://darwin-eu-dev.github.io/CohortSurvival/.

5. Therneau T, Crowson C, Atkinson E. Multi-state models and competing risks. 2024.

6. Catala M, Guo Y, Lopez-Guell K, Burn E, Mercade-Besora N, Alcalde M (2025). *CohortCharacteristics: Summarise and Visualise Characteristics of Patients in the OMOP CDM*. R package version 1.1.0, https://darwin-eu.github.io/CohortCharacteristics/.

7. Recalde M, Manzano-Salgado CB, Díaz Y, Puente D, Garcia-Gil MDM, Marcos-Gragera R, et al. Validation Of Cancer Diagnoses In Electronic Health Records: Results From The Information System For Research In Primary Care (SIDIAP) In Northeast Spain. Clin Epidemiol. 2019 Dec 3;11:1015–24.

8. Ostropolets A, Reich C, Ryan P, Weng C, Molinaro A, DeFalco F, et al. Characterizing database granularity using SNOMED-CT hierarchy. AMIA Annu Symp Proc. 2021 Jan 25;2020:983–92.

9. Campbell WS, Campbell JR, West WW, McClay JC, Hinrichs SH. Semantic analysis of SNOMED CT for a post-coordinated database of histopathology findings. J Am Med Inform Assoc. 2014 Sept 1;21(5):885–92.

11. ANNEXES

ANNEX I. Description of data sources

DATA SOURCES DESCRIPTION

IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)

#	Section	Description
1	Database Identification and country	IQVIA LPD Belgium (IQVIA Longitudinal Patient Database Belgium) Belgium
2	Data partner information section	IQVIA IQVIA Europe
3	Coverage and timespan	Data collection since: 2005 Extent: Nation-wide. Panel of 300 GPs in Belgium. The panel is maintained as a representative sample of the primary care physician population in Belgium, according to three criteria known to influence prescribing: age, sex, and geographical distribution.
4	Healthcare setting / type of data	Primary care – gps. Ambulatory visits, with diagnosis, prescriptions, procedures, and laboratory tests.
5	Data collection process	Outpatient electronic health records. Records are entered by GPs at the healthcare encounter.
6	General representativeness	The panel of contributing physicians (a stable 300 GPs) is maintained as a representative sample of the primary care physician population in Belgium, according to three criteria known to influence prescribing: age, sex, and geographical distribution. The panel consists of a stable 300 GPs that are geographically well spread. The total number of active GPs in Belgium is 15,602. The regional geographical spread of physicians in the LPD data is also representative of the distribution across the country: 57% GPs in the North (compared to 54% nationally), 31% in the South (33% nationally), and 12% in Brussels (13%).The provider of the data has more than 2,250 GPs under contract so in case of a drop out a replacement is easily found.
7	Data content /source coding	No information on source coding.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. The patient ID is per practice. So a patient can have different IDs in the DB, one per practice. In Belgium, patients are typically registered at only one GP practice, so duplication should be minimal.
9	Quality control (database specific)	No QC. Integrity constraints only.
10	Linkage	No linkage.
11	Vital status	Death information is derived from healthcare events.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111116 Website: https://iqvia.com

Danish Data Health Registries (DK-DHR)

#	Section	Description
1	Database Identification and country	DK-DHR (Danish Data Health Registries) Denmark
2	Data partner information section	Danish Medicines Agency (DKMA), Data Analytics Centre (DAC)
3	Coverage and timespan	Data collection since: 1995, Extent: Nationwide. 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 prescription retrievals, vaccination and contraception, Procedures, and Sociodemographic information (sex and age but no information on income, education, occupation).
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. All causes of deaths, all retrieved drug prescriptions, all records of vaccinations, all hospital inpatient and outpatients contacts including disease diagnoses and hospital surgical and non-surgical procedures, histologically confirmed incident cancers, laboratory test results for the entire Danish population from 1/1/1995 onwards.
6	General representativeness	The data is representative of the entire Danish population. Healthcare is free in Denmark, so we do not expect any bias in data collection based on socio-economic status.
7	Data content /source coding	Diagnoses and causes of death are collected using the ICD-10 vocabulary. ATC and RxNorm are used for Drugs. SNOMED codes are used for Procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC, ICDO3, cancer Modifier). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. .
9	Quality control (database specific)	The data we have received relating to nationwide Danish Health Data registries offer an opportunity for large-scale, population-based studies with several advantages 1) Their large size improves the precision of estimates and enables the study of rare exposures and outcomes with long-term latency, 2) Inclusion of nearly all individuals in the target population ensures that the data reflect routine clinical care and all clinical segments of the source population, 3) Data are collected independently of each research study, thus minimising certain types of bias, e.g., non-response, and the influence from attention to the research question on the diagnostic process. Before the source data is sent to us, the Danish Health Data Authority does running and 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.
12	Limitations	There are no clinical measurements in the data. DK-DHR has the following limitations, which may be relevant confounders for certain complex Darwin EU studies: • We lack information on key socio-economic status (SES) factors, such as occupation, education, and income. These variables may be important for analysis in some studies. • We only have complete data on lifestyle factors (such as smoking status and weight) for pregnant women. We have no information on patient contacts in primary care (visits to the GP). Consequently, the incidence of chronic diseases like Type 2 Diabetes might be underestimated.

#	Section	Description
13	Main references	Schmidt M, Schmidt SAJ,Adelborg K,Sundbøll J,Laugesen K,Ehrenstein V,Sørensen HT "The Danish health care system and epidemiological research: from health care contacts to database records." Clinical epidemiology (2019): 31372058
14	Link to HMA-EMA catalogue and database webpage	Website: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/healthdatadenmark HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111217

Estonian Biobank (EBB)

#	Section	Description
1	Database Identification and country	EBB (Estonian Biobank) Estonia
2	Data partner information section	University of Tartu Institute of Computer Science
3	Coverage and timespan	Data collection since: 2004 Extent: Nation-wide. EBB is a nation-wide database containing records from 2004 onwards. Estonian population-based cohort size of 211,800 participants (01/01/2024) aged 18 years and older recruited at GP offices, private practices, and hospitals or in the recruitment offices of the Estonian Genome Center.
4	Healthcare setting / type of data	Primary care – GPs, and community pharmacists, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Registry which collects electronic records from the biobank and cohort study.
5	Data collection process	Data is retrieved by Estonian Biobank once a year from national registries. The insurance claims are requested from Estonian Health Insurance Fund. The inpatient and outpatient electronic health records are requested from National Health Information System. The cancer registry and cause of death registry information is requested from The National Institute for Health Development. The data is sent to the national registry by the healthcare providers.
6	General representativeness	The age, sex, and geographical distribution closely reflect those of the Estonian adult population and encompass close to 5approximately 20% of adult population. Female participants are over-represented in EBB. Overall, 3.4% of Estonian men and 5.5% of Estonian women are represented in EBB. Older people tend to participate less frequently, however, all age groups are well represented.
7	Data content /source coding	All participants have undergone a standardized health assessment, including provision of blood samples for purification of DNA, white blood cells, and plasma, and completed a questionnaire covering various health-related topics, such as lifestyle, diet, and clinical diagnoses. Diseases and health problems are recorded as ICD-10 codes and prescribed medicine according to the ATC classification and local package codes. Procedures and services are coded with NOMESCO classifier and local service codes.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. There is one national identifier that allows linking together all encounters across databases.
9	Quality control (database specific)	The quality control procedures in the Estonian Biobank aim to remove the most obvious mistakes in the data, misspellings, impossible dates, duplicates. Before performing the ETL, several problems are fixed on the source data. Since the ETL procedures are used for a number of different datasets (from the same national sources), we have a growing number of pre-processing steps that correspond to the issues we have discovered previously in the data, such as checking for the presence of critical values,

#	Section	Description
		harmonizing date and unit of measurement formats, checking the validity of certain entries against classifiers, etc.
10	Linkage	Follow-up data are available via linkage with national health-related registries and via re-examination of participants. Furthermore, electronic health records are updated for phenotypic outcome information every year. The EBB database is regularly linked with national registries, hospital databases, and the databases of the Estonian Health Insurance Fund (EHIF) and the National Health Information System (NHIS)
11	Vital status	Vital status (death date and causes of death) are obtained from the Causes of Death Registry.
12	Limitations	Participation in EBB cohort is voluntary, therefore the biobank does not represent a random sample and could be subject to recruitment bias. Although recruitment was open to everyone, there is a disproportion between ethnic Estonians and ethnic Russians in the biobank, with Estonians being overrepresented.
13	Main references	Milani, L., Alver, M., Laur, S. *et al.* The Estonian Biobank's journey from biobanking to personalized medicine. *Nat Commun* **16**, 3270 (2025). https://doi.org/10.1038/s41467-025-58465-3
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111114 Website: https://genomics.ut.ee/en/content/estonian-biobank

Finnish Care Register for Health Care (FinOMOP-THL)

#	Section	Description
1	Database Identification and country	FinOMOP-THL (Finnish Care Register for Health Care) Finland
2	Data partner information section	Finnish Institute for Health and Welfare (THL) Department of Knowledge Brokers
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. The current CDM population comprises all persons having been alive and residing in Finland since the beginning of 2011.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The THL database covers both public and private, primary, and specialised inpatient and outpatient health care encounters in Finland, starting from 2011. The entire public sector and private inpatient encounters have been included since 2011, while private outpatient encounters, including occupational care, are included since 2020. Since 1998, the register has covered both public outpatient and inpatient specialized care and private inpatient care (TerveysHilmo). Since 2009, the Finnish National Vaccination Register is covered (complete since 2020). The vaccination register covers all vaccinations from the public sector and from a large part of private vaccination providers, with the data coverage from both sections being very good from 2020 onwards. Since 2011, the register has covered public primary care (AvoHilmo). Since 2020, the register has covered private outpatient care and occupational care. In addition, the CDM also contains positive COVID-19 test results from the Finnish National Infectious Diseases Register, which is maintained by THL.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. Data is entered by clinicians upon healthcare contact and processed by THL.

#	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, ATC, Toimenpideluokitus (procedure classification adapted from the Nordic Classification of Surgical Procedures (NCSP)), Terveydenhuollon erikoisalat (Hilmo specific provider speciality), Rokotustapa (AR/YDIN National classification for vaccine administration), Tupakointistatus (AR/YDIN National classification for smoking status). Vaccinations are identified on product level based on batch number, trade name, vaccine title, and ATC-code. This is mapped on brand and type in the OMOP CDM.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Each patient in THL has a unique identifier.
9	Quality control (database specific)	The source data collection undergoes a structural and semantic validation before entry into the source database. Additionally, some coded variables undergo quality assessment against the respective code systems post entry into the database. The source registers are also assessed for completeness and coverage, with the aim of improving future collection in the areas where data is lacking.
10	Linkage	THL is already a linkage of multiple Finnish registries (see above).
11	Vital status	The National Population registry data forms the basis for forming the patient population. This ensures an up-to-date location (municipality of residence) of patients, as well as complete death occurrences (although not the cause of death).
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Häkkinen, Pirjo; Mölläri, Kaisa; Saukkonen, Sanna-Mari; Väyrynen, Riikka; Mielikäinen, Lasse; Järvelin, Jutta "Hilmo - Sosiaali- ja terveydenhuollon hoitoilmoitus 2020 : Määrittelyt ja ohjeistus : Voimassa 1.1.2020 alkaen" Terveyden ja hyvinvoinnin laitos (2019):
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111187 Website: https://thl.fi/fi/tilastot-ja-data/ohjeet-tietojen-toimittamiseen/hoitoilmoitusjarjestelma-hilmo

IQVIA Disease Analyzer Germany (IQVIA DA Germany)

#	Section	Description
1	Database Identification and country	IQVIA DA Germany (IQVIA Disease Analyzer Germany) Germany
2	Data partner information section	IQVIA
3	Coverage and timespan	Data collection since: 1989 Extent: Nation-wide. GP and specialists in Germany using specific patient management software.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians). Diagnoses, medication, and procedures from an ambulatory setting. Medications are recorded as prescriptions of marketed products.
5	Data collection process	Outpatient electronic health records. By clinicians at healthcare contact.

#	Section	Description
6	General representativeness	No specific details on general representativeness given.
7	Data content /source coding	Prescription is on product code level (German PZN), ICD10, NFC, Local lab coding.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. There can be patients registered under different ID numbers, because there is no linkage between different GPs.
9	Quality control (database specific)	Data is quality checked on plausibility.
10	Linkage	No.
11	Vital status	Death information is derived from medical events.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104282 Website: https://www.iqvia.com/

Integrated Primary Care Information (IPCI)

#	Section	Description
1	Database Identification and country	IPCI (Integrated Primary Care Information) Netherlands
2	Data partner information section	Erasmus University Medical Center Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards. However, it mainly covers the central part of the country, including the most densely populated area (the 'Randstad') and non-urban areas. IPCI contains information on all patients registered with GPs responsible for non-emergency care and referrals. A patient is registered at birth or at first encounter with the GP.
4	Healthcare setting / type of data	Primary care – gps. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. Data is aggregated by Erasmus MC data managers and combined in one harmonized database. Several checks are done on this database to ensure correct data processing. Persons are mostly uniquely identified, with the exception of when persons change GP practice (when the same individual can receive several different identifiers).
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in

#	Section	Description
		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 v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients can be registered under different IDs. However, in the Netherlands, patients typically have one GP and changing practice is uncommon.
9	Quality control (database specific)	Prior to each data release, extensive quality control steps are performed, e.g., comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g. number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical data, e.g. distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.
13	Main references	de Ridder MAJ, de Wilde M,de Ben C,Leyba AR,Mosseveld BMT,Verhamme KMC,van der Lei J,Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182143
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

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

#	Section	Description
1	Database Identification and country	SIDIAP (The Information System for the Development of Research in Primary Care) Catalunya, Spain
2	Data partner information section	IDIAPJGol
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. The SIDIAP database contains records of around 6 million people residing in Catalonia, estimated to be representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – gps, and hospital inpatient care. SIDIAP captured data includes routine visits, sociodemographic information, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Catala de la Salut is the owner of the data and acts as the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content /source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. No.
9	Quality control (database specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g., there are several variables with information concerning the women's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinally and reliability), plausibility (range checks and unusual values), and consistency are inspected through visualisation tools. In addition, before accessing the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences, and prevalence of the data requested for the project, allowing detection of inconsistencies in the data extraction prior to data delivery. External validation processes of the SIDIAP database mainly include assessing the data

#	Section	Description
		recorded in SIDIAP through linkage to external gold standard data sources, by analysing free text, or by sending questionnaires to health professionals.
10	Linkage	SIDIAP is linked to a hospital discharge database, pharmacy dispensation, and primary care laboratories. It can also be linked to other registries in Catalonia on a project-by-project basis.
11	Vital status	Mortality is fully captured in SIDIAP. The cause of death is not available but can be linked to the Spanish death registry on a project-by-project basis.
12	Limitations	The SIDIAP data is representative of individuals using public primary care. Conditions that are usually followed by specialist care or in private practices might not be properly captured. Patients are followed until death or when transferring to another primary health care centre that does not contribute to SIDIAP.
13	Main references	Recalde M, Rodríguez C,Burn E,Far M,García D,Carrere-Molina J,Benítez M,Moleras A,Pistillo A,Bolíbar B,Aragón M,Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

Clinical Practice Research Datalink GOLD (Oxford) (CPRD GOLD)

#	Section	Description
1	Database Identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) United Kingdom
2	Data partner information section	University of Oxford NDORMS
3	Coverage and timespan	Data collection since: 1987 Extent: Nation-wide. CPRD GOLD consists of patients in contributing practices using Vision software. Historically this covered the whole of the UK, but the number of contributing practices in the England is dropping. In January 2025 only 3 practices from England were a part of CPRD GOLD, while historical patient data were from the whole of the UK, and will continue to be so. In the future, no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records. Data is entered by clinicians into the EHR. Data is processed by CPRD and provides data releases for research.
6	General representativeness	CPRD GOLD has been assessed and found to be broadly representative of the UK general population in terms of age, gender, and ethnicity. In CPRD GOLD in January 2025 there were 2,730,707 current acceptable patients (i.e. registered at currently contributing practices that use Vision software, excluding transferred out, deceased patients, and those flagged by CPRD as not acceptable for clinical research for data quality issues). This equals to 4.07%, based on the UK population estimates of 67,026,300 from the Office of National Statistics (mid-2023). Current patients are only from Scotland, Wales, and Northern Ireland. Historically, GOLD does contain data from England as well.

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

UK BioBank (UKBB)

#	Section	Description
1	Database Identification and country	UKBB (UK BioBank) United Kingdom
2	Data partner information section	Oxford University NDORMS
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. People recruited from whole of the UK.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify).

#	Section	Description
		UK Biobank is made by a rich variety of data sources, which include genetic data, primary care data, hospital inpatient data, death data, and cancer registry.
5	Data collection process	Inpatient hospital electronic health records, and Registries, and Biobank. The baseline assessment is consisting of both patient reported data (questionnaire) and physical measurements. GP and hospital data, as well as death and cancer registry records, are linked afterwards and are subject to data validation: https://biobank.ctsu.ox.ac.uk/~bbdatan/Data_cleaning_overall_doc_showcase_v1.pdf
6	General representativeness	The database population consists of volunteers aged 40-69years. We can expect that volunteers might have been more willing to participate if living closer to one of the 22 recruitment centres or if more interested in health issues compared to the general population. These aspects might have introduced an unavoidable bias in the cohort.
7	Data content /source coding	READ2, READ3, DM+D, ICD9, ICD10, OPCS3, OPCS4, ICD-O-3 are used.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. No.
9	Quality control (database specific)	All UK BioBank data are provided already curated and each of the many datasets have specific curation algorithms and procedures. As always, primary care and hospital data, which come from real-world setting, need special attention regarding data quality. Please refer to the link below for specific details https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/primary_care_data.pdf https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/HospitalEpisodeStatistics.pdf
10	Linkage	The database contains liked data from death and cancer registry, GP and hospital for the participants.
11	Vital status	Linked to the national death registry.
12	Limitations	The UKBB source data will not be updated anymore and have no new records after December 2022. There is no day's supply information captured in the source. GP prescription data are available for 45% of the cohort. There is no information on dispensed medicines. GP laboratory tests and results are not available.
13	Main references	Hewitt J, Walters M,Padmanabhan S,Dawson J "Cohort profile of the UK Biobank: diagnosis and characteristics of cerebrovascular disease." BMJ open (2016): 27006341
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111233 Website: https://www.ukbiobank.ac.uk/

ANNEX II. Fitness for use assessment

Data source justification for inclusion and key characteristics

The selected data sources met the criteria required to capture outcomes of interest and relevant data, enabling a patient-level characterisation of newly diagnosed individuals with cancer across different European settings and regions. The main criterion was a meaningful number of person counts for the population of interest (individuals with cancer) and outcomes (thromboembolic events) assessed at the feasibility stage for all data sources included in the study. Data sources were also selected based on European representativeness. Not all data sources had records of all outcomes of interest.

Additional criteria reflect other data quality domains assessed at the DARWIN EU® data partners onboarding stage. With every new release of the data partners OMOP CDM, the DARWIN EU® coordination centre also receives new results of the *CdmOnboarding*, *DashboardExport,* and *DataQualityDashboard* packages and assesses the quality of the data. No open quality issues related to the study population and outcomes were present for any of the data sources selected for the study.

Relevance was also assessed based on the previous research related to the study population of the outcome. Previously, IPCI, EBB, SIDIAP, CPRD GOLD, IQVIA DA Germany, UKBB, and EBB were used in studies with thromboembolic events as an outcome. (Ali et al., 2020, Mercadé-Besora et al., 2024, Li et al., 2022, Voss et al, 2023). CPRD GOLD, FinOMOP-THL, UKBB, SIDIAP, IPCI, and UKBB were used in studies repeated on individuals with cancer (Hagberg et al., 2023; Corby et al., 2024, Leinonen et al., 2017, Smith et al., 2024; Recalde et al, 2019; van Soest 2008; Chen et al., 2024).

In addition to that, DK-DHR is a nationwide, fully representative data source that includes information from the National Patient Registry and the National Cancer Register

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant
Study population	Objective 1 Inclusion criteria • First diagnosis of a selected cancer (index date) between 01/01/2016 and 31/12/2022 • Age ≥18 years at index date • Minimum 365 days of available history before index date • Index date ≥365 days prior to end of data availability of the data source Exclusion criteria • Diagnoses of multiple primary tumours at index date • History of cancer diagnosis ever before index date Objective 2 Inclusion criteria • Included in the study population of objective 1 • Occurrence of the outcome during follow-up	• First diagnosis of a selected cancer Diagnosis of: • Deep vein thrombosis (DVT) • Pulmonary embolisms (PE) • Venous thromboembolism (VTE, composite of DVT and PE) • Pelvic venous thrombosis (PVT) • Splanchnic vein thrombosis (SVT), including hepatic and extra-hepatic vein thrombosis • Retinal vein thrombosis (RVT), including retinal central vein thrombosis • Disseminated intravascular coagulation (DIC)	Low/Medium/**High**
Treatment/ exposure	Not applicable.		Low/Medium/High
Comparator group (if relevant)	Not applicable.		Low/Medium/High
Outcomes (if relevant)	<u>All objectives</u> The specific thromboembolic events of interest are: • Deep vein thrombosis (DVT) • Pulmonary embolisms (PE)	Diagnosis of: • Deep vein thrombosis (DVT) • Pulmonary embolisms (PE)	Low/Medium/**High**

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant
	• Venous thromboembolism (VTE, composite of DVT and PE) • Pelvic venous thrombosis (PVT) • Splanchnic vein thrombosis (SVT), including hepatic and extra-hepatic vein thrombosis • Retinal vein thrombosis (RVT), including retinal central vein thrombosis • Disseminated intravascular coagulation (DIC) Each of the specific thromboembolic events is a binary outcome. Each outcome is assessed at any diagnosis position in the electronic health record and from any of the care settings in each data source. The list of concepts, from EUPAS1000000440 and based on SNOMED codes and aligned with previous studies that used OMOP CDM and VTE as an outcome (Burn et al., 2022), is provided in **Annex IV**.	• Venous thromboembolism (VTE, composite of DVT and PE) • Pelvic venous thrombosis (PVT) • Splanchnic vein thrombosis (SVT), including hepatic and extra-hepatic vein thrombosis • Retinal vein thrombosis (RVT), including retinal central vein thrombosis • Disseminated intravascular coagulation (DIC)	

Covariates (including confounders if relevant)	All Objectives
	• Sex
	  ○ Female/male
	• Age (years) at date of the first cancer diagnosis, categorised into age groups:
	  ○ 18–34
	  ○ 35–44
	  ○ 45–54
	  ○ 55–64
	  ○ 65–74
	  ○ 75–84
		• Date of first diagnosis of selected cancer (index date)
		• Sex
		• Age at index date
			Low/Medium/**High**

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant
	○ ≥85 Age at index date will be calculated using January 1st of the year of birth as proxy for the actual birthday. Date/month is either not present or cannot be made available for governance reasons. If available, date is often set to first of the month for personal privacy. • Calendar year of the first cancer diagnosis, categorised into subperiods of the study period: ○ 2016–2019 ○ 2020-2022 The status of each covariate will be assessed at index date. Each covariate will be used to stratify the results.		
Follow-up time (if relevant)	For both objectives, follow-up will start on the date of cancer diagnosis (index date) and end on the earliest of occurrence of the outcome, loss to follow-up, end of data availability, death, or end of the study period (31/12/2024).	• Date of first diagnosis of selected cancer (index date) • Date of thromboembolic event • Data of end of data availability • Death date	Low/Medium/**High**

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

ANNEX III. Operational and reporting considerations

DATA MANAGEMENT

Data management

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

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

Data storage and protection

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

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

The output files are stored in the DARWIN EU® Remote Research Environment (RRE). These output files do not contain any data that allow identification of subjects included in the study. The RRE implements further security measures to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/20161 in the various member states.

QUALITY CONTROL

Data source quality control

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

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

The study code will be based on DARWIN EU® R packages: *IncidencePrevalence* to estimate Incidence and Prevalence, *DrugUtilisation* to characterise the drug use, and *CohortCharacteristics* to characterise the cohort by indication. These packages will include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub.

PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

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

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

ANNEX IV. List of stand-alone documents

Concepts to define individuals with cancer (populations of interest) are available in a stand-alone document (DARWIN_EU_P3_C3_005 _Cancer_phenotypes.xlsx). **Tables S1 to S7** include concepts used to define outcomes (thromboembolic events).

Table S1. List of concepts used to define deep vein thrombosis (DVT).

Concept ID	Concept Name	Domain	Vocabulary
762047	Acute bilateral thrombosis of subclavian veins	Condition	SNOMED
762148	Acute deep vein thrombosis of bilateral iliac veins	Condition	SNOMED
37169261	Acute deep vein thrombosis of bilateral lower limbs following procedure	Condition	SNOMED
37169249	Acute deep vein thrombosis of bilateral upper limbs following procedure	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
35615034	Acute deep vein thrombosis of right lower limb following procedure	Condition	SNOMED
35615030	Acute deep vein thrombosis of right upper limb following procedure	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 ileofemoral 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
761461	Acute deep venous thrombosis of bilateral pelvic veins	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 ileofemoral vein of left leg	Condition	SNOMED
765541	Acute deep venous thrombosis of ileofemoral 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

Concept ID	Concept Name	Domain	Vocabulary
761462	Acute deep venous thrombosis of left pelvic vein	Condition	SNOMED
618482	Acute deep venous thrombosis of left peroneal vein	Condition	SNOMED
765537	Acute deep venous thrombosis of left upper extremity	Condition	SNOMED
44782767	Acute deep venous thrombosis of lower extremity as complication of procedure	Condition	SNOMED
44782761	Acute deep venous thrombosis of pelvic vein	Condition	SNOMED
762022	Acute deep venous thrombosis of politeal 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
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
765229	Acute deep venous thrombosis of right pelvic vein	Condition	SNOMED
618681	Acute deep venous thrombosis of right peroneal vein	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
44782766	Acute deep venous thrombosis of upper extremity as complication of procedure	Condition	SNOMED
37171353	Acute ischemia of colon due to thrombosis of mesenteric vein	Condition	SNOMED
37170675	Acute ischemia of small intestine due to thrombosis of mesenteric vein	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
44782762	Acute thrombosis of subclavian vein	Condition	SNOMED
4179911	Axillary vein thrombosis	Condition	SNOMED
37109253	Bilateral acute deep vein thrombosis of femoral veins	Condition	SNOMED
618678	Bilateral acute deep venous thrombosis of peroneal veins	Condition	SNOMED
609003	Bilateral deep femoral vein thrombophlebitis	Condition	SNOMED
3179900	Bilateral deep vein thromboses	Condition	Nebraska Lexicon
40478951	Bilateral deep vein thrombosis of lower extremities	Condition	SNOMED
609002	Bilateral femoral vein thrombophlebitis	Condition	SNOMED
608965	Bilateral iliac vein thrombophlebitis	Condition	SNOMED
1245776	Bilateral popliteal vein thrombophlebitis	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
609006	Bilateral tibial vein thrombophlebitis	Condition	SNOMED
4042396	Deep thrombophlebitis	Condition	SNOMED
4046884	Deep vein thrombosis of leg related to air travel	Condition	SNOMED
3655221	Deep vein thrombosis of lower extremity due to intravenous drug use	Condition	SNOMED
4133004	Deep venous thrombosis	Condition	SNOMED
761013	Deep venous thrombosis of bilateral pelvic veins	Condition	SNOMED
37163011	Deep venous thrombosis of calf	Condition	SNOMED
45773536	Deep venous thrombosis of femoropopliteal vein	Condition	SNOMED
763942	Deep venous thrombosis of left lower extremity	Condition	SNOMED
1075379	Deep venous thrombosis of left posterior tibial vein	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
1075377	Deep venous thrombosis of posterior tibial vein	Condition	SNOMED
4322565	Deep venous thrombosis of profunda femoris vein	Condition	SNOMED
763941	Deep venous thrombosis of right lower extremity	Condition	SNOMED
1075378	Deep venous thrombosis of right posterior tibial vein	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
193512	Embolism and thrombosis of the renal vein	Condition	SNOMED
435565	Embolism and thrombosis of the vena cava	Condition	SNOMED
4258295	Embolism from thrombosis of vein of distal lower extremity	Condition	SNOMED
40481089	Embolism from thrombosis of vein of lower extremity	Condition	SNOMED
40479840	Embolism from thrombosis of vein of thigh	Condition	SNOMED
4119760	Iliofemoral deep vein thrombosis	Condition	SNOMED
4124856	Inferior mesenteric vein thrombosis	Condition	SNOMED
608964	Left iliac vein thrombophlebitis	Condition	SNOMED
602592	Left peroneal vein thrombophlebitis	Condition	SNOMED
600938	Left subclavian vein thrombophlebitis	Condition	SNOMED
37164448	Lemierre syndrome	Condition	SNOMED
4281689	Phlegmasia alba dolens	Condition	SNOMED
4284538	Phlegmasia cerulea dolens	Condition	SNOMED
3185768	Popliteal vein thrombosis	Condition	Nebraska Lexicon
4309333	Postoperative deep vein thrombosis	Condition	SNOMED
1245858	Postpartum acute deep vein thrombosis	Condition	SNOMED
46285905	Provoked deep vein thrombosis	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
608963	Right iliac vein thrombophlebitis	Condition	SNOMED
602583	Right peroneal vein thrombophlebitis	Condition	SNOMED
600939	Right subclavian vein thrombophlebitis	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
4203618	Thrombophlebitis of femoropopliteal vein	Condition	SNOMED
4176614	Thrombophlebitis of iliac vein	Condition	SNOMED
764715	Thrombophlebitis of internal jugular vein	Condition	SNOMED
608904	Thrombophlebitis of left axillary vein	Condition	SNOMED
761821	Thrombophlebitis of left deep femoral vein	Condition	SNOMED
761819	Thrombophlebitis of left femoral vein	Condition	SNOMED
609000	Thrombophlebitis of left popliteal vein	Condition	SNOMED
609005	Thrombophlebitis of left tibial vein	Condition	SNOMED
4318407	Thrombophlebitis of mesenteric vein	Condition	SNOMED
608903	Thrombophlebitis of right axillary vein	Condition	SNOMED
761820	Thrombophlebitis of right deep femoral vein	Condition	SNOMED
761818	Thrombophlebitis of right femoral vein	Condition	SNOMED
609001	Thrombophlebitis of right popliteal vein	Condition	SNOMED
609004	Thrombophlebitis of right tibial vein	Condition	SNOMED
4205652	Thrombophlebitis of subclavian 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

Concept ID	Concept Name	Domain	Vocabulary
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
4153353	Traumatic thrombosis of axillary vein	Condition	SNOMED
46285904	Unprovoked deep vein thrombosis	Condition	SNOMED
37163265	Venous thromboembolism due to thrombosis of vein of lower limb	Condition	SNOMED

Table S2. List of concepts used to define pulmonary embolism (PE).

Concept ID	Concept Name	Domain	Vocabulary
608954	Acute cor pulmonale due to septic pulmonary embolism	Condition	SNOMED
4120091	Acute massive pulmonary embolism	Condition	SNOMED
45768439	Acute pulmonary embolism	Condition	SNOMED
45768888	Acute pulmonary thromboembolism	Condition	SNOMED
762808	Infarction of lung due to embolus	Condition	SNOMED
40480461	Infarction of lung due to iatrogenic pulmonary embolism	Condition	SNOMED
4108681	Postoperative pulmonary embolus	Condition	SNOMED
37160752	Postoperative pulmonary thromboembolism	Condition	SNOMED
1244882	Pulmonary artery embolism due to foreign body	Condition	SNOMED
440417	Pulmonary embolism	Condition	SNOMED
37109911	Pulmonary embolism due to and following acute myocardial infarction	Condition	SNOMED
37016922	Pulmonary embolism on long-term anticoagulation therapy	Condition	SNOMED
43530605	Pulmonary embolism with pulmonary infarction	Condition	SNOMED
4253796	Pulmonary microemboli	Condition	SNOMED
4121618	Pulmonary thromboembolism	Condition	SNOMED
36713113	Saddle embolus of pulmonary artery	Condition	SNOMED
35615055	Saddle embolus of pulmonary artery with acute cor pulmonale	Condition	SNOMED
40479606	Septic pulmonary embolism	Condition	SNOMED
4119607	Subacute massive pulmonary embolism	Condition	SNOMED

Table S3. List of concepts used to define venous thromboembolism (VTE).

Concept ID	Concept Name	Domain	Vocabulary
762047	Acute bilateral thrombosis of subclavian veins	Condition	SNOMED
608954	Acute cor pulmonale due to septic pulmonary embolism	Condition	SNOMED
762148	Acute deep vein thrombosis of bilateral iliac veins	Condition	SNOMED
37169261	Acute deep vein thrombosis of bilateral lower limbs following procedure	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
37169249	Acute deep vein thrombosis of bilateral upper limbs following procedure	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
35615034	Acute deep vein thrombosis of right lower limb following procedure	Condition	SNOMED
35615030	Acute deep vein thrombosis of right upper limb following procedure	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 ileofemoral 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
761461	Acute deep venous thrombosis of bilateral pelvic veins	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 ileofemoral vein of left leg	Condition	SNOMED
765541	Acute deep venous thrombosis of ileofemoral 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
761462	Acute deep venous thrombosis of left pelvic vein	Condition	SNOMED
618482	Acute deep venous thrombosis of left peroneal vein	Condition	SNOMED
765537	Acute deep venous thrombosis of left upper extremity	Condition	SNOMED
44782767	Acute deep venous thrombosis of lower extremity as complication of procedure	Condition	SNOMED
44782761	Acute deep venous thrombosis of pelvic vein	Condition	SNOMED
762022	Acute deep venous thrombosis of politeal vein of right leg	Condition	SNOMED
44782743	Acute deep venous thrombosis of popliteal vein	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
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
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
765229	Acute deep venous thrombosis of right pelvic vein	Condition	SNOMED
618681	Acute deep venous thrombosis of right peroneal vein	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
44782766	Acute deep venous thrombosis of upper extremity as complication of procedure	Condition	SNOMED
37171353	Acute ischemia of colon due to thrombosis of mesenteric vein	Condition	SNOMED
37170675	Acute ischemia of small intestine due to thrombosis of mesenteric vein	Condition	SNOMED
4120091	Acute massive pulmonary embolism	Condition	SNOMED
45768439	Acute pulmonary embolism	Condition	SNOMED
45768888	Acute pulmonary thromboembolism	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
44782762	Acute thrombosis of subclavian vein	Condition	SNOMED
4179911	Axillary vein thrombosis	Condition	SNOMED
37109253	Bilateral acute deep vein thrombosis of femoral veins	Condition	SNOMED
618678	Bilateral acute deep venous thrombosis of peroneal veins	Condition	SNOMED
609003	Bilateral deep femoral vein thrombophlebitis	Condition	SNOMED
3179900	Bilateral deep vein thromboses	Condition	Nebraska Lexicon
40478951	Bilateral deep vein thrombosis of lower extremities	Condition	SNOMED
609002	Bilateral femoral vein thrombophlebitis	Condition	SNOMED
608965	Bilateral iliac vein thrombophlebitis	Condition	SNOMED
1245776	Bilateral popliteal vein thrombophlebitis	Condition	SNOMED
609006	Bilateral tibial vein thrombophlebitis	Condition	SNOMED
44782732	Chronic pulmonary embolism	Condition	SNOMED
45768887	Chronic pulmonary thromboembolism	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
45771016	Chronic pulmonary thromboembolism without pulmonary hypertension	Condition	SNOMED
4042396	Deep thrombophlebitis	Condition	SNOMED
4046884	Deep vein thrombosis of leg related to air travel	Condition	SNOMED
3655221	Deep vein thrombosis of lower extremity due to intravenous drug use	Condition	SNOMED
4133004	Deep venous thrombosis	Condition	SNOMED
761013	Deep venous thrombosis of bilateral pelvic veins	Condition	SNOMED
37163011	Deep venous thrombosis of calf	Condition	SNOMED
45773536	Deep venous thrombosis of femoropopliteal vein	Condition	SNOMED
763942	Deep venous thrombosis of left lower extremity	Condition	SNOMED
1075379	Deep venous thrombosis of left posterior tibial vein	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
1075377	Deep venous thrombosis of posterior tibial vein	Condition	SNOMED
4322565	Deep venous thrombosis of profunda femoris vein	Condition	SNOMED
763941	Deep venous thrombosis of right lower extremity	Condition	SNOMED
1075378	Deep venous thrombosis of right posterior tibial vein	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
193512	Embolism and thrombosis of the renal vein	Condition	SNOMED
435565	Embolism and thrombosis of the vena cava	Condition	SNOMED
4258295	Embolism from thrombosis of vein of distal lower extremity	Condition	SNOMED
40481089	Embolism from thrombosis of vein of lower extremity	Condition	SNOMED
40479840	Embolism from thrombosis of vein of thigh	Condition	SNOMED
4119760	Iliofemoral deep vein thrombosis	Condition	SNOMED
43530934	Induced termination of pregnancy complicated by pulmonary embolism	Condition	SNOMED
762808	Infarction of lung due to embolus	Condition	SNOMED
40480461	Infarction of lung due to iatrogenic pulmonary embolism	Condition	SNOMED
4124856	Inferior mesenteric vein thrombosis	Condition	SNOMED
608964	Left iliac vein thrombophlebitis	Condition	SNOMED
602592	Left peroneal vein thrombophlebitis	Condition	SNOMED
600938	Left subclavian vein thrombophlebitis	Condition	SNOMED
37164448	Lemierre syndrome	Condition	SNOMED
4281689	Phlegmasia alba dolens	Condition	SNOMED
4284538	Phlegmasia cerulea dolens	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
3185768	Popliteal vein thrombosis	Condition	Nebraska Lexicon
4309333	Postoperative deep vein thrombosis	Condition	SNOMED
4108681	Postoperative pulmonary embolus	Condition	SNOMED
37160752	Postoperative pulmonary thromboembolism	Condition	SNOMED
1245858	Postpartum acute deep vein thrombosis	Condition	SNOMED
46285905	Provoked deep vein thrombosis	Condition	SNOMED
1244882	Pulmonary artery embolism due to foreign body	Condition	SNOMED
440417	Pulmonary embolism	Condition	SNOMED
37109911	Pulmonary embolism due to and following acute myocardial infarction	Condition	SNOMED
3655209	Pulmonary embolism due to and following ectopic pregnancy	Condition	SNOMED
3655210	Pulmonary embolism due to and following molar pregnancy	Condition	SNOMED
37016922	Pulmonary embolism on long-term anticoagulation therapy	Condition	SNOMED
43530605	Pulmonary embolism with pulmonary infarction	Condition	SNOMED
4253796	Pulmonary microemboli	Condition	SNOMED
4121618	Pulmonary thromboembolism	Condition	SNOMED
4236271	Recurrent pulmonary embolism	Condition	SNOMED
608963	Right iliac vein thrombophlebitis	Condition	SNOMED
602583	Right peroneal vein thrombophlebitis	Condition	SNOMED
600939	Right subclavian vein thrombophlebitis	Condition	SNOMED
36713113	Saddle embolus of pulmonary artery	Condition	SNOMED
35615055	Saddle embolus of pulmonary artery with acute cor pulmonale	Condition	SNOMED
40479606	Septic pulmonary embolism	Condition	SNOMED
4033521	Splenic vein thrombosis	Condition	SNOMED
4119607	Subacute massive pulmonary embolism	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
4203618	Thrombophlebitis of femoropopliteal vein	Condition	SNOMED
4176614	Thrombophlebitis of iliac vein	Condition	SNOMED
764715	Thrombophlebitis of internal jugular vein	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
608904	Thrombophlebitis of left axillary vein	Condition	SNOMED
761821	Thrombophlebitis of left deep femoral vein	Condition	SNOMED
761819	Thrombophlebitis of left femoral vein	Condition	SNOMED
609000	Thrombophlebitis of left popliteal vein	Condition	SNOMED
609005	Thrombophlebitis of left tibial vein	Condition	SNOMED
4318407	Thrombophlebitis of mesenteric vein	Condition	SNOMED
608903	Thrombophlebitis of right axillary vein	Condition	SNOMED
761820	Thrombophlebitis of right deep femoral vein	Condition	SNOMED
761818	Thrombophlebitis of right femoral vein	Condition	SNOMED
609001	Thrombophlebitis of right popliteal vein	Condition	SNOMED
609004	Thrombophlebitis of right tibial vein	Condition	SNOMED
4205652	Thrombophlebitis of subclavian 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
4153353	Traumatic thrombosis of axillary vein	Condition	SNOMED
46285904	Unprovoked deep vein thrombosis	Condition	SNOMED
37163265	Venous thromboembolism due to thrombosis of vein of lower limb	Condition	SNOMED

Table S4. List of concepts used to define pelvic vein thrombosis (PVT) (concept sets included all descendants of listed concepts).

Concept ID	Concept Name	Domain	Vocabulary
762148	Acute deep vein thrombosis of bilateral iliac veins	Condition	SNOMED
35616028	Acute deep vein thrombosis of left iliac vein	Condition	SNOMED
35616027	Acute deep vein thrombosis of right iliac vein	Condition	SNOMED
765155	Acute deep venous thrombosis of bilateral ileofemoral veins	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
761461	Acute deep venous thrombosis of bilateral pelvic veins	Condition	SNOMED
762015	Acute deep venous thrombosis of ileofemoral vein of left leg	Condition	SNOMED
765541	Acute deep venous thrombosis of ileofemoral vein of right lower extremity	Condition	SNOMED
761462	Acute deep venous thrombosis of left pelvic vein	Condition	SNOMED
44782761	Acute deep venous thrombosis of pelvic vein	Condition	SNOMED
765229	Acute deep venous thrombosis of right pelvic vein	Condition	SNOMED
608965	Bilateral iliac vein thrombophlebitis	Condition	SNOMED
765152	Chronic deep vein thrombosis of bilateral iliac veins	Condition	SNOMED
35616026	Chronic deep vein thrombosis of left iliac vein	Condition	SNOMED
761439	Chronic deep vein thrombosis of left pelvic vein	Condition	SNOMED
46271548	Chronic deep vein thrombosis of pelvic vein	Condition	SNOMED
35616025	Chronic deep vein thrombosis of right iliac vein	Condition	SNOMED
761441	Chronic deep vein thrombosis of right pelvic vein	Condition	SNOMED
765542	Chronic deep venous thrombosis of bilateral ileofemoral veins	Condition	SNOMED
761440	Chronic deep venous thrombosis of bilateral pelvic veins	Condition	SNOMED
765543	Chronic deep venous thrombosis of left ileofemoral vein	Condition	SNOMED
762016	Chronic deep venous thrombosis of right ileofemoral vein	Condition	SNOMED
761013	Deep venous thrombosis of bilateral pelvic veins	Condition	SNOMED
4133975	Deep venous thrombosis of pelvic vein	Condition	SNOMED
608964	Left iliac vein thrombophlebitis	Condition	SNOMED
4285751	Pelvic thrombophlebitis in puerperium	Condition	SNOMED
608963	Right iliac vein thrombophlebitis	Condition	SNOMED
4176614	Thrombophlebitis of iliac vein	Condition	SNOMED
4317290	Thrombophlebitis of pelvic vein	Condition	SNOMED
4111868	Thrombophlebitis of the common iliac vein	Condition	SNOMED
4110343	Thrombophlebitis of the external iliac vein	Condition	SNOMED
4109877	Thrombophlebitis of the internal iliac vein	Condition	SNOMED
42538533	Thrombosis of iliac vein	Condition	SNOMED
4319327	Thrombosis of pelvic vein	Condition	SNOMED

Table S5. List of concepts used to splanchnic vein thrombosis (SVT).

Concept ID	Concept Name	Domain	Vocabulary
37171353	Acute ischemia of colon due to thrombosis of mesenteric vein	Condition	SNOMED
37170675	Acute ischemia of small intestine due to thrombosis of mesenteric vein	Condition	SNOMED
45757410	Acute thrombosis of mesenteric vein	Condition	SNOMED
36712892	Acute thrombosis of splenic vein	Condition	SNOMED
196715	Budd-Chiari syndrome	Condition	SNOMED
4301208	Hepatic vein thrombosis	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
4124856	Inferior mesenteric vein thrombosis	Condition	SNOMED
4092406	Portal thrombophlebitis	Condition	SNOMED
199837	Portal vein thrombosis	Condition	SNOMED
4033521	Splenic vein thrombosis	Condition	SNOMED
4055089	Superior mesenteric vein thrombosis	Condition	SNOMED
4318407	Thrombophlebitis of mesenteric vein	Condition	SNOMED
4317289	Thrombosis of mesenteric vein	Condition	SNOMED

Table S6. List of concepts used to define retinal vein thrombosis (RVT).

Concept ID	Concept Name	Domain	Vocabulary
437544	Arterial retinal branch occlusion	Condition	SNOMED
3657106	Bilateral occlusion of branch retinal arteries	Condition	SNOMED
37310623	Bilateral occlusion of central retinal arteries	Condition	SNOMED
37169454	Bilateral vascular occlusion of retina of eyes	Condition	SNOMED
4336004	Branch macular artery occlusion	Condition	SNOMED
4339013	Branch retinal vein occlusion with macular edema	Condition	SNOMED
4334248	Branch retinal vein occlusion with neovascularization	Condition	SNOMED
4199035	Branch retinal vein occlusion with no neovascularization	Condition	SNOMED
437540	Central retinal artery occlusion	Condition	SNOMED
313761	Central retinal vein occlusion	Condition	SNOMED
4208221	Central retinal vein occlusion - ischemic	Condition	SNOMED
4208222	Central retinal vein occlusion - non-ischemic	Condition	SNOMED
4339010	Central retinal vein occlusion with macular edema	Condition	SNOMED
4334246	Central retinal vein occlusion with neovascularization	Condition	SNOMED
4338905	Cilioretinal artery occlusion	Condition	SNOMED
42535735	Combined occlusion by thrombus of retinal artery and retinal vein	Condition	SNOMED
4102317	Incipient occlusion of retinal vein	Condition	SNOMED
4083482	Macular branch retinal vein occlusion	Condition	SNOMED
37206377	Occlusion of branch of retinal vein of left eye	Condition	SNOMED
37206378	Occlusion of branch of retinal vein of right eye	Condition	SNOMED
37206381	Occlusion of central retinal vein of left eye	Condition	SNOMED
37206380	Occlusion of central retinal vein of right eye	Condition	SNOMED
36713329	Occlusion of left branch retinal artery	Condition	SNOMED
37207955	Occlusion of left central retinal artery	Condition	SNOMED
3657873	Occlusion of left cilioretinal artery	Condition	SNOMED
36713330	Occlusion of right branch retinal artery	Condition	SNOMED
37207895	Occlusion of right central retinal artery	Condition	SNOMED
3657872	Occlusion of right cilioretinal artery	Condition	SNOMED

Concept ID	Concept Name	Domain	Vocabulary
4334245	Retinal artery occlusion	Condition	SNOMED
4324290	Retinal phlebitis	Condition	SNOMED
440392	Retinal vascular occlusion	Condition	SNOMED
3183076	Right branch retinal artery occlusion	Condition	Nebraska Lexicon
4216561	Thrombophlebitis of retinal vein	Condition	SNOMED
4187790	Thrombosis of retinal vein	Condition	SNOMED
3657847	Vascular occlusion of retina of left eye	Condition	SNOMED
3657848	Vascular occlusion of retina of right eye	Condition	SNOMED
312622	Venous retinal branch occlusion	Condition	SNOMED

Table S7. List of concepts used to define disseminated intravascular coagulation (DIC).

Concept ID	Concept Name	Domain	Vocabulary
37117819	Acquired purpura fulminans	Condition	SNOMED
436093	Disseminated intravascular coagulation	Condition	SNOMED
4028488	Purpura fulminans	Condition	SNOMED

ANNEX V. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Doc.Ref. EMA/540136/2009

Adopted by the ENCePP Steering Group on 15/10/2018

Study title: DARWIN EU® - Time to onset of thromboembolic events in adults with selected types of cancer

EU PAS Register® number: EUPAS1000000814
Study reference number: P4-C2-017

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]	X	☐	☐	8.5
1.1.2 End of data collection[2]	X	☐	☐	8.5
1.1.3 Progress report(s)	X	☐	☐	5
1.1.4 Interim report(s)	X	☐	☐	5
1.1.5 Registration in the EU PAS Register®	X	☐	☐	5
1.1.6 Final report of study results.	X	☐	☐	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:				
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)	X	☐	☐	6
2.1.2 The objective(s) of the study?	X	☐	☐	7
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	X	☐	☐	8.3
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	X	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	X	

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)	X	☐	☐	8.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	X	☐	☐	8.4
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	X	☐	☐	8.8.3
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))	☐	☐	X	
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)	☐	☐	X	

Comments:

Section 4: Source and study populations		**Yes**	**No**	**N/A**	**Section Number**
4.1	Is the source population described?	X	☐	☐	Annex I
4.2	Is the planned study population defined in terms of:				
	4.2.1 Study time period	X	☐	☐	8.5
	4.2.2 Age and sex	X	☐	☐	8.3
	4.2.3 Country of origin	X	☐	☐	8.4
	4.2.4 Disease/indication	X	☐	☐	8.3
	4.2.5 Duration of follow-up	X	☐	☐	8.2
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	X	☐	☐	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)	☐	☐	X	
5.2	Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☐	X	
5.3	Is exposure categorised according to time windows?	☐	☐	X	

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

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	X	☐	☐	8.6.1
6.2 Does the protocol describe how the outcomes are defined and measured?	X	☐	☐	8.6.1
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)	X	☐	☐	9
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)	☐	☐	X	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☐	☐	X	
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	X	☐	☐	9
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	X	☐	☐	9

Comments:

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

Comments:

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☐	☐	X	
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	X	☐	☐	Annex I
9.1.3 Covariates and other characteristics?	X	☐	☐	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)	☐	☐	X	
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	X	☐	☐	Annex I
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	X	☐	☐	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)	☐	☐	X	
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	X	☐	☐	Annex I
9.3.3 Covariates and other characteristics?	X	☐	☐	Annex I
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	X	

Comments:

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

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

Comments:

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

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	X	☐	☐	9
12.1.2 Information bias?	X	☐	☐	9
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	X	☐	☐	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)	X	☐	☐	8.7

Comments:

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	X	☐	☐	Annex III

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	X	
13.3 Have data protection requirements been described?	X	☐	☐	Annex III

Comments:

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

Comments:

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

Comments:

ANNEX VI. Glossary

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

Aggregated Data

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

Benefit-Risk Assessment

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

Common Data Model (CDM)

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

Complex Studies (C3)

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

Coordination Centre (CC)

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

Data Access

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

Data Quality Framework

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

Data Source

A database 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 databases 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 individual 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 databases, 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.

Name of the main author of the protocol: Melissa Leung

Date: 10/12/2025

Signature: M. Leung