NATURAL HISTORY OF COAGULOPATHY AND USE OF ANTI-THROMBOTIC AGENTS IN COVID-19 PATIENTS AND PERSONS VACCINATED AGAINST SARS-COV-2 – REPORT 2: OBSERVED VS EXPECTED RATES OF EVENTS OF INTEREST FOLLOWING VACCINATION AND INFECTION WITH SARS-COV-2

Main authors

Daniel Prieto-Alhambra, MD MSc PhD
Department of Medical Informatics
EMC – Rotterdam
Centre for Statistics in Medicine (CSM)
NDORMS – University of Oxford

Edward Burn, PhD
Centre for Statistics in Medicine (CSM)
NDORMS – University of Oxford
SIDIAP-Idiap Jordi Gol, Universitat Autonoma de Barcelona

Document Status: Version 1.0 -

Date: 29th June 2021

EU PAS register number: EUPAS40414

PASS information

Title	Natural history of coagulopathy and use of anti-thrombotic agents in COVID-19 patients and persons vaccinated against SARS-CoV-2
Protocol version identifier	1.0
Date of last version of protocol	12/04/2021
EU PAS register number	EUPAS40414
Active Ingredient	n/a
Medicinal product	J07BX
Product reference	n/a
Procedure number	n/a
Marketing authorisation holder(s)	n/a
Joint PASS	n/a
Research question and objectives	1) To estimate the background incidence of selected embolic and thrombotic events of interest among the general population. 2) To estimate the incidence of selected embolic and thrombotic events of interest among persons vaccinated against SARS-CoV-2 at 7, 14, 21, and 28 days. 3) To estimate incidence rate ratios for selected embolic/thrombotic events of interest amongst people vaccinated against SARS-CoV-2 compared to background rates as estimated in Objective #1. 4) To estimate the incidence of venous thromboembolic events among patients with COVID-19 at 30-, 60-, and 90-days. 5) To calculate the risks of COVID-19 worsening stratified by the occurrence of a venous thromboembolic event. 6) To assess the impact of risk factors on the rates of venous thromboembolic events among patients with COVID-19.

	7) To develop and externally validate patient-level prediction models for venous thromboembolic events for patients with COVID-19. 8) To estimate the incidence of arterial thromboembolic events among patients with COVID-19 at 30-, 60-, and 90-days. 9)To calculate the risks of COVID-19 worsening stratified by the occurrence of an arterial thromboembolic event. 10) To assess the impact of risk factors on the rates of arterial thromboembolic events among patients with COVID-19. 11) To develop and externally validate patient-level prediction models for arterial thromboembolic events for patients with COVID-19. **The current progress report contains results for Objectives 2 and 3, and partial results for Objective 4.**
Country(-ies) of study	Italy, France, Germany, Netherlands, Spain, and United Kingdom. **The current report contains data form Spain and the UK.**
Author	Edward Burn (University of Oxford, Idiap Jordi Gol) Daniel Prieto-Alhambra (University of Oxford, EMC Rotterdam)

Marketing authorisation holder(s)

Marketing authorisation holder(s)	Not applicable
MAH contact person	Not applicable

1. Table of contents

1. Table of contents 4
2 List of abbreviations 6
3 Responsible parties 7
4 Abstract 8
5 Amendments and updates 12
6 Milestones 13
7 Rationale and background 14
7.1 Coagulopathy in the general population and among individuals vaccinated against SARS-CoV-2 14
7.2 Sources of heterogeneity between contributing databases 14
8 The two latter were illustrated and documented by studying detailed codes for DVT and their impact on background rates across our network.Research question and objectives 16
9 Research methods 17
9.1 Study design 17
9.2 Setting 17
9.2.1 Countries 17
9.2.2 Study period 17
9.2.3 Study cohorts 17
9.2.4 Follow-up 17
9.3 Variables 17
9.3.1 Study outcomes 17
9.3.2 Characteristics of study participants 19
9.4 Data sources 21
9.5 Study size 23
9.6 Data management 23
9.7 Data analysis 23
9.7.1 Analytical code 23
9.7.2 Descriptive statistics 23
9.7.3 Background incidence rates 23
9.7.4 Sources of heterogeneity across databases 24
9.8 Quality control 24
9.8.1 General database quality control 24
9.8.2 Study-specific quality control 24

9.9 Limitations of the research methods 25
10 Protection of human subjects 26
11 Management and reporting of adverse events/adverse reactions 27
12 Plans for disseminating and communicating study results 28
13 Study Results 29
13.1 Baseline Characteristics 29
13.2 Incidence rates of thromboembolic events 31
13.3 Incidence rates of coagulopathies and thrombocytopenia 36
13.4 Database heterogeneity 40
13.4.1 Differences in coding practice across data sources 40
13.4.2 Impact of data linkage 41
13.5 TTS patient characterisation 42
14 Conclusions 45
15 References 47

2 List of abbreviations

Abbreviation	Name
ATC	Anatomical Therapeutic Chemical Classification
CDM	Common Data Model
COVID-19	Coronavirus disease-2019
CPRD	Clinical Practice Research Datalink
DA	Disease Analyzer
ECMO	Extracorporeal membrane oxygenation
EHR	Electronic Health Record
HES APC	Hospital Episode Statistics Admitted Patient Care
HM	Hospital de Madrid
IPCI	Integrated Primary Care Information
LPD	Longitudinal Patient Data
MACE	Major cardiovascular events
OMOP	Observational Medical Outcomes Partnership
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SIDIAP	The Information System for Research in Primary Care

3 Responsible parties

Senior Project Manager

Prof Johan van der Lei (EMC)

Project Management

Associate Prof Peter Rijnbeek (EMC)

Prof Dani Prieto-Alhambra (University of Oxford)

Principal Investigators

Prof Dani Prieto-Alhambra (University of Oxford)

Associate Prof Katia Verhamme (EMC)

Associate Prof Peter Rijnbeek (EMC)

Clinical Epidemiologists

Prof Dani Prieto-Alhambra (University of Oxford)

Associate Prof Katia Verhamme (EMC)

Dr. Talita Duarte-Salles (IDIAPJGol)

Statistician/s

Dr. Victoria Strauss (University of Oxford)

Ms. Xintong Li (University of Oxford)

Dr. Maria de Ridder (EMC)

Data controllers/analysts

Marcel de Wilde (EMC)

Mees Mosseveld (EMC)

Kristin Kostka (IQVIA)

Edward Burn (University of Oxford and IDIAPJGol)

Database Stewards

Prof Johan van der Lei (IPCI)

Dr. Christian Reich (IQVIA)

Dr. Talita Duarte-Salles (SIDIAP, HM)

Dr. Antonella Delmestri (CPRD GOLD and AURUM)

4 Abstract

Title

Thromboembolic events and thrombosis with thrombocytopenia after COVID-19 infection and vaccination in Catalonia (Spain) and the UK

Version and Date: Version 1.0, 29th June 2021

Name and affiliation of main author: Daniel Prieto-Alhambra (University of Oxford, Erasmus MC)

Rationale and background:

Venous (VTE) and arterial thromboembolism (VTE), alone and with concurrent thrombocytopenia, are being investigated in association with some COVID-19 vaccines.

Research question and objectives

The current report contains results on one objective as part of a larger study. The objective covered here is Objective 1) To estimate the background incidence of selected embolic and thrombotic events of interest among the general population.

Study design

We performed a European international network cohort study using data from 6 European countries: France, Germany, Italy, Netherlands, Spain, and the United Kingdom (UK).

Population

The current report focuses on historical rates in the source population registered in each of the contributing data sources in the period between 2017 and 2019. Sensitivity analyses were conducted focused on 1) people with at least one year of data visibility before index date; and 2) people with at least one healthcare visit after 1/1/2017.

Variables

Patient demographics, health conditions and medication/s use were extracted.

Outcomes of interest included venous thromboembolic events (deep vein thrombosis DVT, pulmonary embolism PE, cranial vein thrombosis CVT, visceral venous thrombosis) alone and in combination with concomitant thrombocytopenia (i.e. thrombosis-thrombocytopenia syndromes TTS), arterial thromboembolic events, coagulopathy events, cardiovascular events, and mortality.

Data sources

Primary care records from Netherlands (IPCI), Italy (IQVIA LPD Italy), Spain (SIDIAP), and the UK (CPRD GOLD); ambulatory data from France (IQVIA LPD France) and Germany (IQVIA DA Germany). a subset of SIDIAP (primary care) was further linked to inpatient data from Spain (from here on reported as SIDIAP-H) to explore the impact of such linkage on completeness and on the resulting background rate estimates. All contributing data sources were mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model.

Study size

All the individuals satisfying the eligibility criteria and registered in contributing databases were included. This included the following number/s of people: >1.2 million (IPCI NL), >1.1 million (IQVIA LPD IT), >5.7 million (SIDIAP ES) of which >1.9 million in SIDIAP-H, >3.9 million (CPRD UK), >8.4 million (IQVIA DA DE), and >3.9 million (IQVIA LPD FR).

Data analyses

Incidence rates of study outcomes in the general population (i.e. background rates) were estimated per 100,000 person-years for two periods: 1) 1/1/2017 to 31/12/2019, and 2) date of a first visit in 2017-2019 until 31/12/2019. Rates were estimated stratified by database, age, and sex.

Database heterogeneity was explored in terms of baseline socio-demographics, contributing source and mapped (standard) codes, and by comparing SIDIAP and SIDIAP-H findings to further understand the impact of linkage to hospital records.

People with TTS were characterised in terms of age, sex, and previous comorbidity and use of medicines in the previous 6 months.

All analyses were conducted in a distributed network fashion, and all analytical code is available here: https://github.com/oxford-pharmacoepi/CovCoagBackgroundIncidence

Results

All study results, including characterisation and background rates of all events and sensitivity analyses, are reported in full in an interactive website: https://livedataoxford.shinyapps.io/CovCoagBackgroundIncidence/

Rates of DVT ranged from 86/100,000 (DE) to 187/100,000 (NL), and increased with older age in all databases. Rates of DVT with thrombocytopenia were much lower and went from 1/100,000 (NL) to 1.5 (IT), and increased with age. Rates of PE ranged from 66/100,000 (IT) to 92/100,000 (NL), and higher in older populations. Rates of PE with thrombocytopenia were again lower, from 0.5/100,000 (FR) to 1.5/100,000 (DE).

CVT and splanchnic vein thrombosis were less common, from 0.3/100,000 (FR) to 1.2/100,000 (ES, UK) and from 1.5/100,000 (FR) to 14.6/100,000 (ES) respectively. CVT with thrombocytopenia was only observed in SIDIAP-H, at a rate of 0.1/100,000. Splanchnic vein thrombosis with thrombocytopenia was seen in UK, DE and SIDIAP-H ES, with a similar rate of 0.1/100,000 person-years in the two former, and of 0.7/100,000 in the latter.

Regarding coagulopathies, thrombocytopenia was common, with rates 168/100,000 (FR) to 523/100,000 (UK). Conversely, immune thrombocytopenia (IT), heparin-induced thrombocytopenia (HIT), and disseminated intravascular coagulation (DIC) were rare, all with rates <20/100,000.

Background rates of other events are reported in Appendix 1 and in the aforementioned web application.

Differences in coding practice were illustrated by the use of unspecific codes for DVT in LPD France: the three most commonly used ICD10 codes for DVT in this database were unspecific for deep veins and therefore excluded from DVT narrow (but included for DVT broad). Coverage/linkage also had an apparent impact on background rates: we observed substantial increases in background rates when SIDIAP was linked to hospital records (SIDIAP-H). E.g. CVST rates jumped from 0.1 to 1.2/100,000 in SIDIAP vs SIDIAP-H respectively.

Finally, we found patients with TTS to be generally older and in worse health than the source population and compared to those with thromboses without thrombocytopenia. E.g. in CPRD, the source population had a prevalence of autoimmune disease of 1.8%, compared to 4.2% for DVT (narrow) and 12.6% for DVT with thrombocytopenia. Similarly, the prevalence of cancer was 5.1%, 15.6% and 25.2% respectively, and the use of NSAIDs and systemic corticosteroids jumped from 23.0% to 53.8% and 62.2% and from 10.3% to 24.9% and 29.9% respectively.

Conclusions

There is heterogeneity in the background rates of the studied coagulopathy, thromboembolic and TTS events in terms of age-sex and geography/database. Where possible, a same population and data source should be studied to analyse post-vaccine vs background rates for comparison.

The proportion of thromboembolic events with concomitant thrombocytopenia (TTS) is <1/50 for DVT/PE, but higher (about 1/10-1/20) for CVT.

Although thrombocytopenia and platelet disorders are relatively common (background rates of up to >500/100,000 person-years), the other studied coagulopathies are rare events, all with rates <20/100,000.

Key factors contributing to background rate heterogeneity across databases included differences in coding practice, and coverage/linkage to hospital records.

Patients with TTS were typically older, predominantly of male gender, and had a higher prevalence of comorbidity and medicines use than those with thrombosis without thrombocytopenia (and much higher than the source population). Research is needed to clarify whether historically observed TTS is equivalent to post-vaccination TTS.

5 Amendments and updates

There have been no formal amendments to the protocol so far.

Number	Date	Section of study protocol	Amendment or update	Reason
1				
2				
…				

6 Milestones

Milestone	Planned date
Approval Study Protocol by EMA	March 2021
<Registration in the EU PAS register>	*March 2021*
Start of data collection	*February 2021*
End of data collection	*May 2021*
Draft report	July 2021
Final study report accepted by EMA	August 2021
Manuscript to be provided to EMA	September 2021

7 Rationale and background

7.1 Coagulopathy in the general population and among individuals vaccinated against SARS-CoV-2

Various vaccines for coronavirus disease-2019 (COVID-19) have been authorised by regulators including the European Medicines Agency, the Food and Drug Administration in the United States, and the United Kingdom Medicines and Healthcare products Regulatory Agency following phase 3 clinical efficacy trials. Millions of individuals have since received one of these vaccines. As with all medical products, however, there remains a need to continue to monitor safety.

Routinely collected health care data can provide valuable evidence as to the incidence of various events of interests among people who have been vaccinated. Moreover, such data sources can be used to estimate the incidence of such events among the general population to contextualise the findings. One particular area of interest are embolic and thrombotic events. At the time of writing (30th March, 2021) case reports of a small number of serious blood clots among individuals vaccinated with the Astra-Zeneca COVID-19 vaccine have been reported. Further investigation into this specific safety concern is warranted. Using routinely collected data to assess the incidence of such events of interest among the general population and in individuals vaccinated against SARS-CoV-2 would provide valuable evidence in assessing this specific safety concern.

The current report focuses on the background incidence rates of thromboembolic events, thrombosis-thrombocytopenia, and coagulopathy events in the general population of 6 European countries: DE, ES, FR, IT, NL, and the UK.

7.2 Sources of heterogeneity between contributing databases

Although not pre-specified per protocol, we explored sources of heterogeneity between contributing databases at the request of EMA reviewers. Three potential causes exist for database heterogeneity:

1. **True differences in the epidemiology of a disease in the source population** due to geographic, ethnic, socio-demographic, or genetic factors.
2. **Differences in coding practice** due to differences in source coding system, differences (usually subtle) between country-specific coding systems, and other differences related to cultural, language-related or healthcare differences. For example, this can be seen in the literature by the impact on research findings when databases switched from using ICD-9 to ICD-10 codes (link).

3. **Differences in database coverage** due to differences in healthcare systems (eg general practitioners as gatekeepers in the UK or Spain but not in Germany or France) and in linkage availability (eg linkage to hospital records).

8 The two latter were illustrated and documented by studying detailed codes for DVT and their impact on background rates across our network.Research question and objectives

The objective covered in the current progress report is as follows:

1) **To estimate the background incidence of selected embolic and thrombotic events of interest among the general population.**

In addition, we explored sources of database heterogeneity as outlined above (7.1).

Other objectives will follow in subsequent reports, namely:

2) To estimate the incidence of selected embolic and thrombotic events of interest among persons vaccinated against SARS-CoV-2 at 7, 14, 21, and 28 days.
3) To estimate incidence rate ratios for selected embolic/thrombotic events of interest amongst people vaccinated against SARS-CoV-2 compared to background rates as estimated in Objective #1
4) To estimate the incidence of venous thromboembolic events among patients with COVID-19 at 30, 60, and 90 days.
5) To calculate the risks of COVID-19 worsening stratified by the occurrence of a venous thromboembolic event.
6) To assess the impact of risk factors on the rates of venous thromboembolic events among patients with COVID-19.
7) To develop and externally validate patient-level prediction models for venous thromboembolic events for patients with COVID-19.
8) To estimate the incidence of arterial thromboembolic events among patients with COVID-19 at 30-, 60-, and 90-days.
9) To calculate the risks of COVID-19 worsening stratified by the occurrence of an arterial thromboembolic event.
10) To assess the impact of risk factors on the rates of arterial thromboembolic events among patients with COVID-19.
11) To develop and externally validate patient-level prediction models for arterial thromboembolic events for patients with COVID-19.

9 Research methods

9.1 Study design

An observational cohort study using routinely-collected health care data mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

9.2 Setting

9.2.1 Countries

Datasets from Italy, France, Germany, Netherlands, Spain, and United Kingdom will inform the analyses (see section 9.4 Data Sources below for more details).

9.2.2 Study period

The study period for estimating the background incidence of events of interest starts from the 1st January 2017 and end on 31st December 2019.

9.2.3 Study cohorts

The following study cohorts were defined:

1. **General population cohorts (date anchored)**
    - present in the database as of the 1st January 2017 (1st January will be the index date)
2. **General population cohorts (visit anchored)**
    - with a visit/contact with the healthcare system between 1st January 2017 and 31st December 2019 (with the date of that first visit used as index date)

Both cohorts were generated with and without imposing a 1-year presence in the database to illustrate the impact of this inclusion criterion. The analyses without the 1-year criterion are reported in Appendix 1.

9.2.4 Follow-up

Follow-up went from cohort-specific index date and until the first of: outcome of interest, loss to follow-up, or end of the study period.

9.3 Variables

9.3.1 Study outcomes

Full concept sets for all study outcomes are available in the attached Appendix 2.

9.3.1.1 Venous thromboembolic events

In a primary analysis, *venous thromboembolic events* (VTE) were identified by diagnostic codes for pulmonary embolism or deep vein thrombosis. In a secondary analysis pulmonary embolism (PE) and deep vein thrombosis (DVT) were assessed separately.

9.3.1.2 Arterial thromboembolic events

In the primary analysis, arterial thromboembolic events were identified by an acute myocardial infarction or acute ischemic stroke. In a secondary analysis acute myocardial infarction and acute ischemic stroke will be assessed separately.

9.3.1.3 Other thromboembolic events

We identified the occurrence of the following outcomes identified as potential safety signals associated with some COVID-19 vaccines:

1. Cerebral venous thrombosis
2. Splenic vein thrombosis
3. Splenic artery thrombosis
4. Splenic infarction
5. Hepatic vein thrombosis
6. Hepatic artery thrombosis
7. Portal vein thrombosis
8. Intestinal infarction
9. Mesenteric vein thrombosis
10. Celiac artery thrombosis
11. Visceral vein thrombosis

9.3.1.4 Coagulopathy and thrombocytopenia

We ascertained the occurrence of the following events identified as potential safety signals associated with some COVID-19 vaccines:

1. Disseminated intravascular coagulation
2. Immune thrombocytopenia
3. Thrombotic thrombocytopenia purpura
4. Heparin-induced thrombocytopenia
5. Thrombocytopenia
6. Thrombocytopenic purpura
7. Platelet disorder/s

9.3.1.5 Thrombosis-thrombocytopenia syndromes (TTS)

We looked at TTS as the co-occurrence of each of the thromboembolic events in 9.3.1.1 and 9.3.1.3 in combination with thrombocytopenia. Co-occurrence was defined as thrombocytopenia identified/recorded within 10 days before/after the diagnosis of the thrombotic event.

9.3.1.6 Cardiovascular events

Instances of heart failure, cardiac arrhythmia, ventricular arrhythmia or cardiac arrest, chest pain or angina, and sudden cardiac death will be identified. In addition, major cardiovascular events (MACE) will be identified by heart failure, acute myocardial infarction, or stroke, or the occurrence of sudden cardiac death. As a sensitivity analysis, we will require that events were observed during a hospitalisation.

9.3.2 Characteristics of study participants

9.3.2.1 Demographics

Patients´ age at index date and sex will be identified.

9.3.2.2 Health conditions and medications use pre-index date

Prevalence of key health conditions as recorded any time before index date are reported for the entire source populations in Table 1. Similarly, medications use as recorded in the six months before index date are reported for the source populations in Table 1.
The list of selected conditions and medicines is as follows:

• Autoimmune disease
• Antiphospholipid syndrome
• Thrombophilia
• Asthma
• Atrial fibrillation
• Malignant neoplastic disease
• Diabetes mellitus

• Obesity
• Heart disease
• Hypertensive disorder
• Renal impairment
• COPD
• Dementia

The list of selected medicines used in the 6-month period before index and reported here includes:

• Non-steroidal anti-inflammatory drugs
• Cox2 inhibitors
• Systemic corticosteroids
• Antithrombotic and anticoagulant therapies
• Lipid modifying agents
• Antineoplastic and immunomodulating agents
• Hormonal contraceptives for systemic use
• Tamoxifen
• Sex hormones and modulators of the genital system

A range of other health conditions and medicines used were identified in the 30 days and 365 days before index date for the source population and for each of the outcome cohorts, and are reported in full in an interactive web application: see 'Patient profiles' tab here.

9.4 Data sources

For this study, we used routinely-collected healthcare data from databases throughout Europe. These databases are summarised in Table A below. All of these databases had been mapped to the OMOP CDM. Feasibility analyses (described below in section 9.8.2) were run prior to executing the analysis. Based on this, we added the subset of SIDIAP with linked inpatient data to maximise completeness in the ascertainment of many of the study outcomes. Both are reported to illustrate the impact of hospital data linkage.

Table A: Data sources accessible for analysis

Database	Managing Organization	Country	Description
LPD Italy	IQVIA	Italy	LPD Italy is comprised of anonymised patient records collected from software used by GPs during an office visit to document patients´ clinical records. Data coverage includes over 2M patient records historically with at least one visit and 119.5M prescription orders across 900 GP practices. Dates of service include from 2004 through present. Observation time is defined by the first and last consultation dates. Drugs are captured as prescription records with product, quantity, dosing directions, strength, indication and date of consultation.
LPD France	IQVIA	France	LPD France is a computerised network of physicians including GPs who contribute to a centralised database of anonymised patient EMR. Currently, >1200 GPs from 400 practices are contributing to the database covering historically >7.8M patients in France. The database covers a time period from 1994 through the present. Observation time is defined by the first and last consultation dates. Drug information is derived from GP prescriptions. Drugs obtained over the counter by the patient outside the prescription system are not reported.
DA Germany	IQVIA	Germany	IQVIA DA Germany is collected from extracts of patient management software used by GPs and specialists practicing in ambulatory care settings. Data coverage includes more than 34M distinct person records historically, collected from 2,734

			providers. Dates of service include from 1992 through March 2020.
CPRD GOLD	UOXF	UK	The Clinical Practice Research Datalink (CPRD) is a governmental, not-for-profit research service, jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA), a part of the Department of Health, United Kingdom (UK). CPRD GOLD contains data contributed by GP practices using Vision® software.
IPCI	Erasmus MC	Netherlands	The Integrated Primary Care Information (IPCI) database is collected from EHR records of patients registered with 391 GPs throughout the Netherlands. The database contains records from approximately 2.6 million patients historically starting in 1996.
SIDIAP	IDIAP Jordi Gol	Spain	The Information System for Research in Primary Care (SIDIAP; www.sidiap.org) is a primary care records database that covers approximately 80% of the population of Catalonia, North-East Spain. Healthcare is universal and tax-payer funded in the region, and primary care physicians are gatekeepers for all care and responsible for repeat prescriptions.
SIDIAP-H	IDIAP Jordi Gol	Spain	SIDIAP-H is a subset of about 30% of SIDIAP (about 2 million people) which contains linked administrative data related to hospital admissions equivalent to the Conjunt Minim Basic de Dades a l'Alta Hospitalaria (CMBD-AH) for hospitals pertaining to the Catalan Institute of Health.

It is worth noting that HM Hospitals was not included as this database only includes COVID-19 affected patients, and will be therefore used for subsequent analyses. Similarly, CPRD AURUM does not provide background population data, but just extracts of study cohorts, precluding the analysis of background incidence rates in the general population. Linked HES is not available for the general population, again precluding its use for background rates estimation in the analysis of CPRD GOLD or AURUM. Finally, SIDIAP and SIDIAP-H were analysed separately to inform the added value of hospital inpatient linkage.

9.5 Study size

For each database, all individuals that satisfied the eligibility criteria for a study cohort were included. The number of people eligible in the primary analysis are reported in Table 1.

9.6 Data management

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

The analytic code for this study was written in R. Each data partner executed the study code against their database containing patient-level data and then returned the results set which only contained aggregated data. The results from each of the contributing data sites was finally combined in tables and figures for the study report.

9.7 Data analysis

9.7.1 Analytical code

All analytical code, cohort definitions, and programming documentation are available in an open repository to maximise transparency and reproducibility: https://github.com/oxford-pharmacoepi/CovCoagBackgroundIncidence

9.7.2 Descriptive statistics

The observed characteristics of each study population as a whole are reported in the current report. Additional information, and outcome-specific cohorts are also characterised in our interactive web application, under the 'Patient profile' tab:
https://livedataoxford.shinyapps.io/CovCoagBackgroundIncidence/

9.7.3 Background incidence rates

We estimated the incidence for all of the study outcomes described in section 9.3.2 among the general population study cohorts. Incidence rates were estimated as the total number of events divided by the person-time at risk per 100,000 person-years, with 95% confidence intervals provided. As well as estimating the incidence of outcomes of interest in general, we estimated them stratified by age (three groupings to align with FDA Sentinel protocol, EMA requirements, and wider categories for better precision when reporting on rare outcomes) and sex. The analyses reported in this document are those based on the general population with 1+ years of data visibility. Analyses without this requirement, and rates estimated after an "anchoring" healthcare visit are reported in Appendix 1 and in the accompanying interactive app.

9.7.4 Sources of heterogeneity across databases

Medical vocabularies vary and the databases used in the study use Read codes (CPRD), ICPC (IPCI), ICD-9 (LPD Italy), ICD-10 (DA Germany and LPDD France), and ICD-10CM (SIDIAP) to represent condition-related concepts. These coding systems differ in the way that they describe clinical events (in particular their granularity) and this can have a meaningful impact on research findings. Differences in coding practice were illustrated by exploring the impact of different DVT codes on DVT broad and DVT narrow rates in the contributing databases.

We demonstrated the impact of linkage by comparing the estimated background rates of DVT in SIDIAP and SIDIAP-H.

9.8 Quality control

9.8.1 General database quality control

A number of open-source quality control mechanisms for the OMOP CDM have been developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). In particular, it is expected that data partners will have run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard) . This tool provides a number of checks relating to the conformance, completeness and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions, completeness in the sense of data quality is solely focused on quantifying missingness, or the absence of data, while plausibility seeks to determine the believability or truthfulness of data values.[21] Each of these categories has one or more subcategories and are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions.

9.8.2 Study-specific quality control

Each of the contributing databases run a cohort diagnostics package (https://github.com/oxford-pharmacoepi/CovCoagOutcomeDiagnostics) to identify the outcome cohorts described above. The results of this exercise are available in an interactive web application, and include the following elements based on the identification of subjects with each of the study outcomes in the whole database history:

1. Cohort counts: N of records and patients with each of the recorded outcomes

2. Incidence rate: rough estimates of incidence rates for each of the outcomes, stratified by age, sex, index year, and database
3. Time distributions: descriptive of time (in days) available before index date
4. Included concepts: list of standard as well as source codes included for each outcome and database
5. Orphan concepts: list of potential codes that could resemble the study outcomes
6. Index event breakdown: count of subjects included based on each of the included concepts
7. Visit context: healthcare setting where a given event was identified in each database, where this is available (only SIDIAP-H)
8. Cohort characterization: deep characterisation of all recorded diagnoses and medicines use any time before index date for each event cohort
9. Temporal characterization: similar to the above but using pre-specified time windows (year or month before/after index date)
10. Cohort overlap: depiction of the number (%) of patients with each of the study outcomes, and their overlap in the entire database history
11. Compare cohort characteristics: characterization of more than one cohort for comparison
12. Database information: brief description of each of the contributing databases

Each of these elements were reviewed by the PI of the project (Prof D Prieto-Alhambra) and one other study member (Dr Edward Burn), and iterations of this used to identify missing (orphan) codes, irrelevant/implausible ones, secular trends and consistency of recording over time, and overlap between study outcome cohorts.
Illustrative examples of differences in coding practices that contribute to database heterogeneity are highlighted in this report, namely DVT broad and DVT narrow in LPD France compared to the other contributing data sources.

9.9 Limitations of the research methods

The study was informed by routinely-collected health care data and so data quality issues must be considered. The included databases vary in the data elements that they capture. Not all outcomes were available in all databases, given the context where some of the rarest events are treated and diagnosed. For example, CVT was not be observed in all databases.

10 Protection of human subjects

For this study, participants from various EU member states processed personal data from individuals which is collected in national/regional electronic health record databases. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy. All the databases used in this study are already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses were run, which generate non-identifiable aggregate summary results. Where required, Institutional Review Boards of the respective databases and ethics committees reviewed the protocol of the study.

Regulatory and ethical compliance

This study was designed and implemented and reported in accordance with the Guidelines for Good Pharmacoepidemiology Practices (GPP) of the International Society for Pharmacoepidemiology, the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines, and with the ethical principles laid down in the Declaration of Helsinki.

This study is fulfilling the criteria of a ‘European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) study’ and follows the ‘ENCePP Code of Conduct’.

11 Management and reporting of adverse events/adverse reactions

According to the new guideline on good pharmacovigilance practice (EMA/873138/2011 Rev 2*) there is no requirement for expedited reporting of adverse drug reactions from studies with secondary use of data (such as electronic health care databases).

12 Plans for disseminating and communicating study results

Dissemination activities to be undertaken will have mainly, although not exclusively, a scientific nature (articles, presentations at conferences, etc.). In addition to the current report, a manuscript is in preparation, focussing on the more relevant aspects of our analyses.

13 Study Results

13.1 Baseline Characteristics

The number of subjects available in background populations was (in order) as follows: 1.1 million in LPD IT, 1.3 million in IPCI-NL, 5.8 million in SIDIAP ES, 1.9 million in SIDIAP-H ES, 3.9 million in CPRD-UK, 4.0 in LPD France, and 8.5 million in DA Germany.

There was a predominance of female participants in all databases, and median (IQR) age ranged from 41 (22-59) in CPRD-UK to 52 (37-68) years old in LPD Italy. All contributing data sources had a long follow-up available before 2017, with a median (IQR) ranging from 3.2 (1.8-5.7) to 11.9 (4.7-15.1) in IPCI-NL and CPRD-UK respectively. More detailed baseline characteristics for study participants in the 7 contributing databases are reported in Table 1.

SIDIAP and SIDIAP-H were comparable in terms of age, sex, and prior observation time, but the latter had slightly higher prevalence of specific comorbidities e.g. atrial fibrillation (2.1% in SIDIAP vs 3.1% in SIDIAP-H), cancer (5.4% vs 7.4%), and slightly higher previous use of some medicines (eg systemic corticosteroids) and lower of others (eg NSAIDs).

Table 1. Baseline characteristics of source study population/s

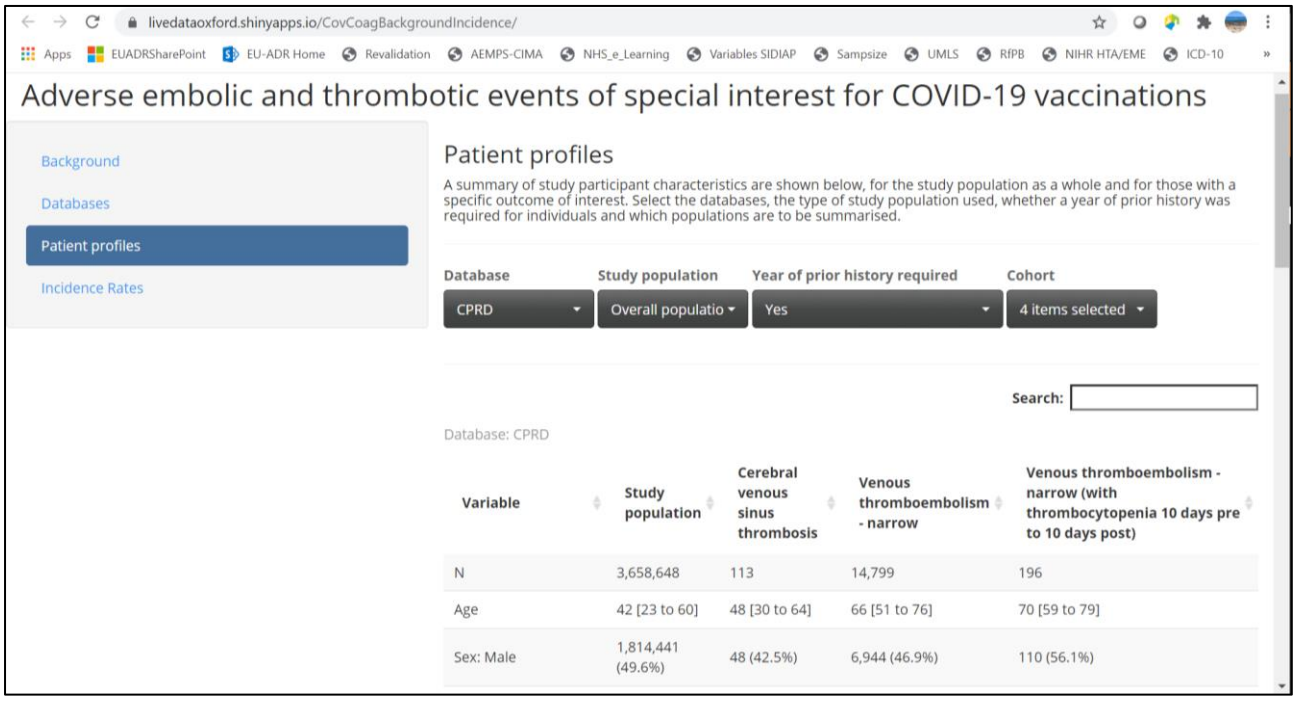
	CPRD	**DA Germany**	**LPD France**	**IPCI**	**LPD Italy**	**SIDIAP**	**SIDIAP_H**
N	3,913,071	8,459,098	3,951,633	1,299,288	1,066,230	5,779,691	1,909,814
Age (Median [IQR])	41 [22 to 59]	52 [32 to 67]	48 [28 to 65]	44 [23 to 60]	52 [37 to 68]	42 [25 to 59]	43 [26 to 61]
Sex: Male	1,937,858 (49.5%)	3,589,506 (42.4%)	1,669,415 (42.2%)	636,386 (49.0%)	426,758 (40.0%)	2,850,570 (49.3%)	945,873 (49.5%)
Years of prior observation time (Median [IQR])	11.9 [4.7 to 15.1]	4.8 [1.9 to 8.9]	4.6 [2.0 to 6.2]	3.2 [1.8 to 5.7]	6.3 [5.0 to 6.5]	11.0 [11.0 to 11.0]	11.0 [11.0 to 11.0]
Comorbidities							
Autoimmune disease	70,604 (1.8%)	238,985 (2.8%)	32,245 (0.8%)	24,645 (1.9%)	45,567 (4.3%)	74,482 (1.3%)	29,599 (1.5%)
Antiphospholipid syndrome	1,166 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	901 (0.0%)	93 (0.0%)
Thrombophilia	3,039 (0.1%)	6,474 (0.1%)	313 (0.0%)	0 (0.0%)	0 (0.0%)	920 (0.0%)	97 (0.0%)
Asthma	484,991 (12.4%)	412,789 (4.9%)	222,161 (5.6%)	138,777 (10.7%)	79,528 (7.5%)	335,739 (5.8%)	114,573 (6.0%)

Atrial fibrillation	76,091 (1.9%)	92,767 (1.1%)	13,412 (0.3%)	31,801 (2.4%)	34,325 (3.2%)	120,665 (2.1%)	59,505 (3.1%)
Malignant neoplastic disease	198,275 (5.1%)	534,352 (6.3%)	66,962 (1.7%)	106,223 (8.2%)	86,645 (8.1%)	313,793 (5.4%)	141,336 (7.4%)
Diabetes mellitus	213,996 (5.5%)	597,233 (7.1%)	174,564 (4.4%)	93,035 (7.2%)	95,611 (9.0%)	451,848 (7.8%)	185,752 (9.7%)
Obesity	107,522 (2.7%)	530,958 (6.3%)	15,634 (0.4%)	40,395 (3.1%)	46,101 (4.3%)	904,262 (15.6%)	340,786 (17.8%)
Heart disease	278,323 (7.1%)	936,730 (11.1%)	194,630 (4.9%)	129,562 (10.0%)	165,172 (15.5%)	541,483 (9.4%)	239,735 (12.6%)
Hypertensive disorder	558,671 (14.3%)	1,425,782 (16.9%)	498,244 (12.6%)	222,433 (17.1%)	322,776 (30.3%)	1,090,196 (18.9%)	429,924 (22.5%)
Renal impairment	168,610 (4.3%)	169,166 (2.0%)	13,064 (0.3%)	27,555 (2.1%)	31,853 (3.0%)	195,479 (3.4%)	87,587 (4.6%)
COPD	80,393 (2.1%)	358,047 (4.2%)	41,040 (1.0%)	40,116 (3.1%)	27,119 (2.5%)	134,704 (2.3%)	61,492 (3.2%)
Dementia	33,537 (0.9%)	95,957 (1.1%)	9,217 (0.2%)	7,873 (0.6%)	10,458 (1.0%)	65,929 (1.1%)	26,587 (1.4%)
Medication use (183 days prior to four days prior)							
Non-steroidal anti-inflammatory drugs	900,092 (23.0%)	928,497 (11.0%)	1,056,021 (26.7%)	211,464 (16.3%)	293,188 (27.5%)	1,615,456 (28.0%)	322,184 (16.9%)
Cox2 inhibitors	7,126 (0.2%)	33,006 (0.4%)	13,769 (0.3%)	8,165 (0.6%)	21,899 (2.1%)	27,042 (0.5%)	8,253 (0.4%)
Systemic corticosteroids	404,443 (10.3%)	269,020 (3.2%)	315,054 (8.0%)	139,482 (10.7%)	84,587 (7.9%)	337,036 (5.8%)	120,338 (6.3%)
Antithrombotic and anticoagulant therapies	114,246 (2.9%)	213,378 (2.5%)	175,535 (4.4%)	44,985 (3.5%)	110,079 (10.3%)	112,852 (2.0%)	46,510 (2.4%)
Lipid modifying agents	143,424 (3.7%)	191,407 (2.3%)	197,031 (5.0%)	54,039 (4.2%)	83,234 (7.8%)	81,698 (1.4%)	38,573 (2.0%)
Antineoplastic and immunomodulating agents	124,080 (3.2%)	210,390 (2.5%)	94,702 (2.4%)	54,941 (4.2%)	36,792 (3.5%)	64,047 (1.1%)	10,509 (0.6%)
Hormonal contraceptives for systemic use	173,708 (4.4%)	169,549 (2.0%)	98,852 (2.5%)	47,983 (3.7%)	18,740 (1.8%)	46,817 (0.8%)	11,513 (0.6%)
Tamoxifen	2,141 (0.1%)	3,761 (0.0%)	826 (0.0%)	865 (0.1%)	684 (0.1%)	1,230 (0.0%)	478 (0.0%)

Sex hormones and modulators of the genital system	213,023 (5.4%)	228,846 (2.7%)	141,501 (3.6%)	55,810 (4.3%)	29,750 (2.8%)	58,958 (1.0%)	18,866 (1.0%)

Baseline characteristics for people affected with each of the events of interest are available from our online interactive application: https://livedataoxford.shinyapps.io/CovCoagBackgroundIncidence/ under the 'Patient profiles' tab (see screenshot in Figure 1).

Figure 1. Screenshot of interactive web application reporting patient characteristics for an illustrative example of background population, CVST, VTE, and VTE with thrombocytopenia in CPRD-UK data



13.2 Incidence rates of thromboembolic events

Incidence rates of all outcomes of interest are reported in detail in Appendix 1 in the form of a spreadsheet, stratified by database, sex, and age. Appendix 1 includes estimates of background rates obtained from sensitivity analyses, including those restricted to people with a 1-year data visibility requirement, and those obtained after an encounter with the healthcare system. Neither of both sensitivity analyses resulted in major changes in the estimated rates.

Overall unadjusted rates (and 95% confidence intervals) of thromboembolic events of interest (CVT, DVT, PE, and visceral venous thromboses), alone and with thrombocytopenia, are reported in full in Table 2, stratified by database (excluding SIDIAP, reported in 13.4).

Table 2. IRs of thromboembolism and TTS (per 100,000 PYs)

	N	PYs	Number of events	Incidence rate per 100, 000 PYs (95% confidence interval)
Cerebral venous sinus thrombosis				
CPRD	3,913,025	9,676,085	118	1.2 (1.0 to 1.5)
DA Germany	8,459,044	19,369,671	95	0.5 (0.4 to 0.6)
France LPD	3,951,606	8,210,128	26	0.3 (0.2 to 0.5)
SIDIAP_H	1,909,812	5,491,342	65	1.2 (0.9 to 1.5)
Cerebral venous sinus thrombosis (with thrombocytopenia 10 days pre to 10 days post)				
SIDIAP_H	1,909,814	5,491,404	6	0.1 (0.0 to 0.2)
Deep vein thrombosis – narrow definition				
CPRD	3,909,649	9,656,721	9,071	93.9 (92.0 to 95.9)
DA Germany	8,451,032	19,329,175	16,600	85.9 (84.6 to 87.2)
IPCI	1,296,310	3,402,027	6,367	187.2 (182.6 to 191.8)
Italy LPD	1,066,587	2,639,975	3,891	147.4 (142.8 to 152.1)
SIDIAP_H	1,908,025	5,479,231	5,854	106.8 (104.1 to 109.6)
Deep vein thrombosis narrow (with thrombocytopenia 10 days pre to 10 days post)				
CPRD	3,913,031	9,676,132	127	1.3 (1.1 to 1.6)
DA Germany	8,458,995	19,369,390	225	1.2 (1.0 to 1.3)

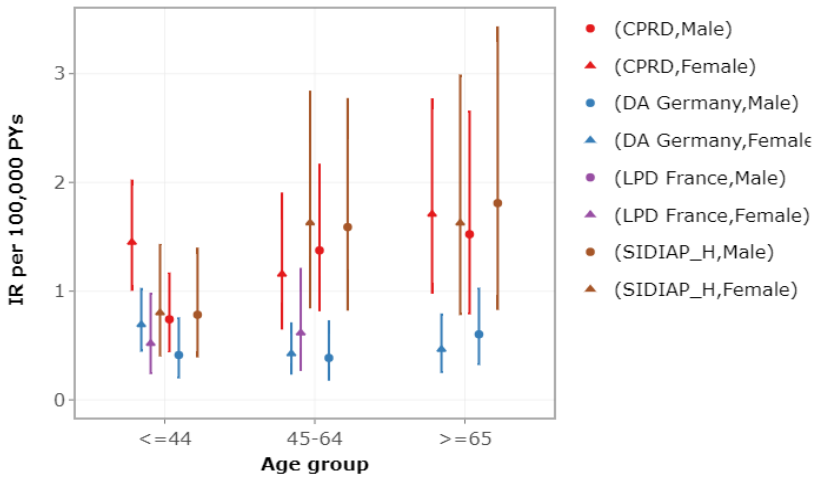
IPCI	1,299,274	3,418,833	34	1.0 (0.7 to 1.4)
Italy LPD	1,066,209	2,651,714	39	1.5 (1.0 to 2.0)
SIDIAP_H	1,909,800	5,491,326	57	1.0 (0.8 to 1.3)
Pulmonary embolism (narrow)				
CPRD	3,910,531	9,662,585	7,149	74.0 (72.3 to 75.7)
DA Germany	8,449,246	19,325,560	17,204	89.0 (87.7 to 90.4)
France LPD	3,947,450	8,195,137	4,700	57.4 (55.7 to 59.0)
IPCI	1,297,807	3,410,984	3,141	92.1 (88.9 to 95.4)
Italy LPD	1,064,532	2,645,601	1,748	66.1 (63.01 to 69.2)
SIDIAP_H	1,908,546	5,483,310	4,225	77.1 (74.7 to 79.4)
Pulmonary embolism (with thrombocytopenia 10 days pre to 10 days post)				
CPRD	3,913,042	9,676,256	84	0.9 (0.7 to 1.1)
DA Germany	8,458,971	19,369,265	286	1.5 (1.3 to 1.7)
France LPD	3,951,605	8,210,109	39	0.5 (0.3 to 0.6)
IPCI	1,299,282	3,418,860	21	0.6 (0.4 to 0.9)
Italy LPD	1,066,222	2,651,761	17	0.6 (0.4 to 1.0)
SIDIAP_H	1,909,806	5,491,321	79	1.4 (1.1 to 1.8)
Splanchnic Vein Thrombosis				
CPRD	3,913,005	9,675,960	233	2.4 (2.1 to 2.7)
DA Germany	8,458,941	19,369,177	398	2.1 (1.9 to 2.3)

France LPD	3,951,594	8,210,016	122	1.5 (1.2 to 1.8)
Italy LPD	1,066,207	2,651,684	58	2.2 (1.7 to 2.8)
SIDIAP_H	1,909,669	5,490,342	799	14.6 (13.6 to 15.6)
Splanchnic Vein Thrombosis (with thrombocytopenia 10 days pre to 10 days post)				
CPRD	3,913,070	9,676,375	5	0.1 (0.0 to 0.1)
DA Germany	8,459,086	19,369,887	16	0.1 (0.0 to 0.1)
SIDIAP_H	1,909,808	5,491,364	36	0.7 (0.5 to 0.9)

The overall, unadjusted rates of CVT [95% confidence interval] ranged from 0.3/100,000 person-years [0.2-0.5] in LPD FR to 1.2/100,000 [0.9-1.5] and 1.2 [1.0-1.5] in SIDIAP-H ES and CPRD UK respectively. As demonstrated in Figure 2, rates were numerically higher in older people and similar in both genders. Substantial uncertainty demonstrated in the form of confidence intervals precluded comparison between age and sex groups.

CVST with concomitant thrombocytopenia was only estimable in SIDIAP-H ES, and had an overall unadjusted incidence rate of 0.1/100,000 py [0.0 to 0.2]. The limited number of events observed limited our ability to estimate age or sex-specific rates for this particular outcome.

Figure 2. Age and gender-specific incidence rates of CVST

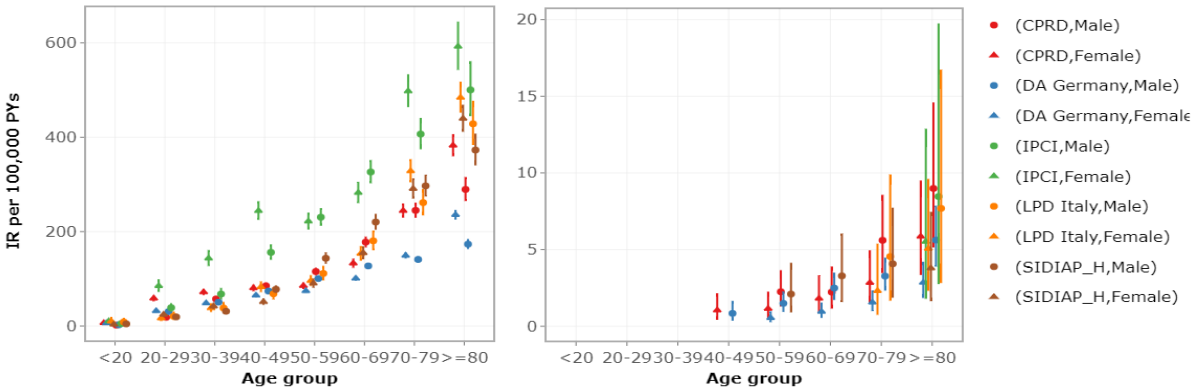


The unadjusted overall background rates of DVT (in its narrow/more specific definition) ranged between 85.9 [84.6 to 87.2] in DA Germany and 187.2 [182.6 to 191.8] in IPCI NL. DVT with

concomitant thrombocytopenia was about 100-fold less common, with rates ranging between 1.0 [0.7 to 1.4] in IPCI NL (and similar figures in SIDIAP-H ES) to 1.5 [1.0 to 2.0] in LPD IT. Both events increased with age, as shown in Figure 3. DVT was more common in women in all ages.

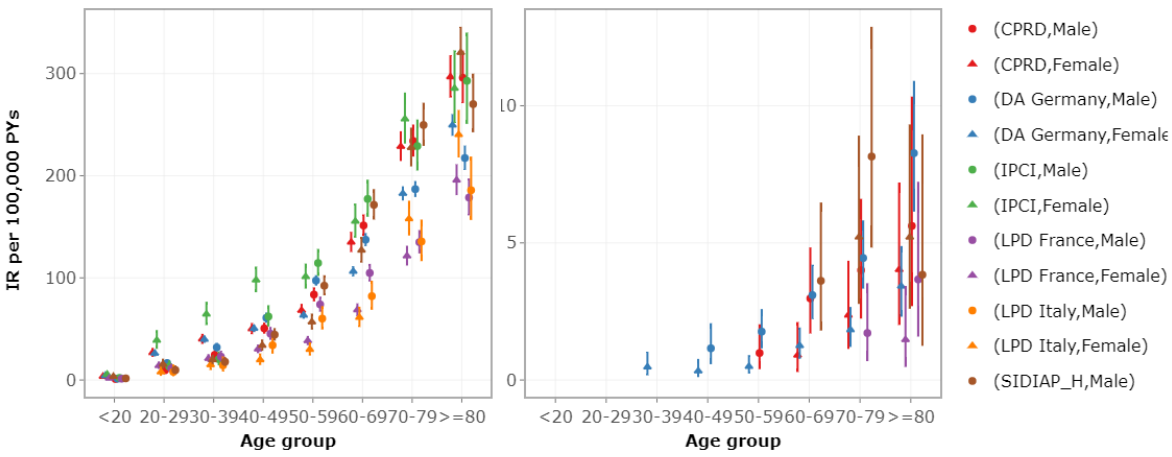
It is worth noticing that no rates of DVT are reported for France, as the used codes were deemed too unspecific to be included in the narrow (specific) algorithm for the identification of DVT events.

Figure 3. Age and gender-specific background rates of DVT (left) and DVT with associated thrombocytopenia (right)



Unadjusted rates of PE ranged from 66.1/100,000 py [63.0 to 69.2] in LPD IT to 92.1 [88.9 to 95.4] in IPCI NL. PE with thrombocytopenia was again much lower, and ranged from 0.5/100,000 py [0.3 to 0.6] in LPD FR to 1.5 [1.3 to 1.7] in DA DE. As shown in Figure 4, both events increased with older age. Gender differences were more obvious in younger ages, with women having higher rates of PE (but not clearly for PE TTS) in ages <50 years old.

Figure 4. Age and gender-specific background rates of PE (left) and PE with associated thrombocytopenia (right)



Finally, splanchnic vein thrombosis had an overall unadjusted background rate ranging between 1.5 (1.2 to 1.8) in LPD FR to 14.6 (13.6 to 15.6) per 100,000 py in SIDIAP-H ES. The second highest was 2.4 (2.1 to 2.7) in CPRD UK, possibly demonstrating the impact of hospital linkage for this particular outcome. Splanchnic vein thrombosis with thrombocytopenia was only estimable in CPRD and SIDIAP-H ES, with background rates of 0.1/100,000 py in both cases.

13.3 Incidence rates of coagulopathies and thrombocytopenia

Background rates for all the platelet disorders and coagulopathies of interest, stratified by age, sex, and data source are reported in Appendix 1. A summary of the ones of most recent interest is reported in Table 3.

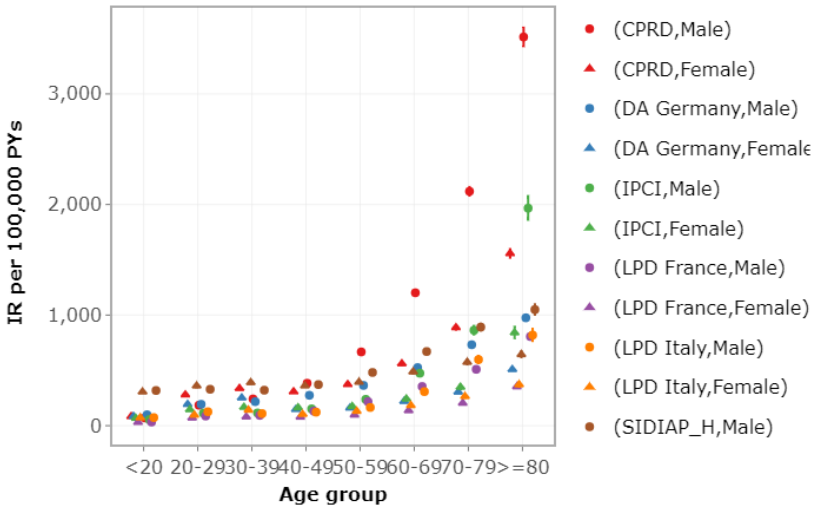
Table 3. IRs of coagulopathies (per 100,000 PYs)

	N	PYs	Number of events	Incidence rate per 100, 000 PYs (95% confidence interval)
Thrombocytopenia				
CPRD	3,872,432	9,510,358	49,752	523.1 (518.5 to 527.8)
DA Germany	8,417,342	19,183,971	57,195	298.1 (295.7 to 300.6)
France LPD	3,942,984	8,172,936	13,721	167.9 (165.0 to 170.7)
IPCI	1,295,622	3,398,241	8,532	251.1 (245.8 to 256.5)
Italy LPD	1,062,466	2,635,277	5,524	209.6 (204.1 to 215.2)
SIDIAP_H	1,896,893	5,419,118	24,010	443.1 (437.5 to 448.7)
Heparin-induced thrombocytopenia				
CPRD	3,912,943	9,675,822	302	3.1 (2.8 to 3.5)
DA Germany	8,458,456	19,366,573	1513	7.8 (7.4 to 8.2)
France LPD	3,951,623	8,210,192	20	2.8 (2.6 to 3.1)
Italy LPD	1,066,144	2,651,425	171	6.4 (5.5 to 7.5)

SIDIAP_H	1,909,715	5,490,687	396	7.2 (6.5 to 8.0)
Disseminated intravascular coagulation				
CPRD	3,913,067	9,676,361	15	0.2 (0.1 to 0.3)
DA Germany	8,459,041	19,369,706	79	0.4 (0.3 to 0.5)
France LPD	3,951,611	8,210,148	34	0.4 (0.3 to 0.6)
Italy LPD	1,066,206	2,651,697	37	1.4 (1.0 to 1.9)
SIDIAP_H	1,909,755	5,491,127	206	3.8 (3.3 to 4.3)
Immune thrombocytopenia				
CPRD	3,912,708	9,674,616	759	7.8 (7.3 to 8.4)
DA Germany	8,457,949	19,364,321	2,264	11.7 (11.2 to 12.2)
France LPD	3,951,527	8,209,807	175	2.1 (1.8 to 2.5)
IPCI	1,299,133	3,418,075	267	7.8 (6.9 to 8.8)
SIDIAP_H	1,909,599	5,489,765	918	16.7 (15.7 to 17.8)

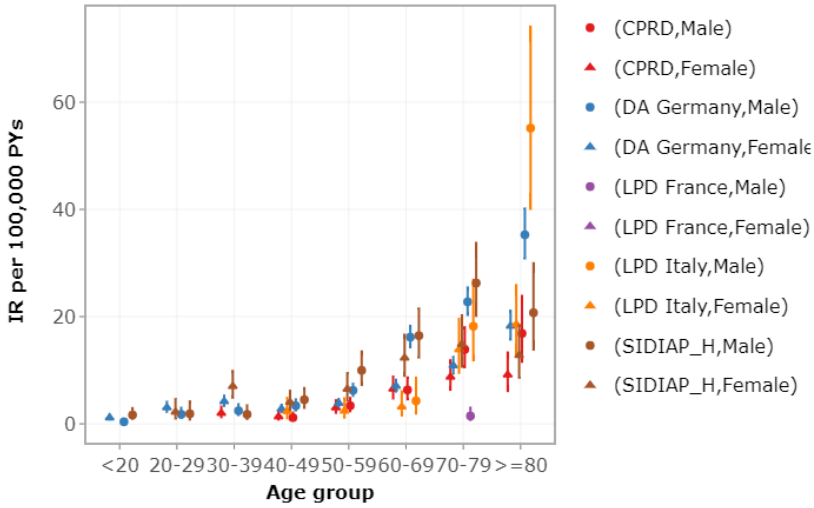
Thrombocytopenia was by far the most common syndrome, and (based on cohort diagnostics) relied almost entirely on laboratory data availability. The observed unadjusted background rates of this event ranged from 167.9 (165.0 to 170.7) in LPD FR to 523.1 (518.5 to 527.8) in CPRD UK. The incidence of thrombocytopenia increased with age, and was higher in men than women as demonstrated in Figure 5.

Figure 5. Age and gender-specific background rates of thrombocytopenia



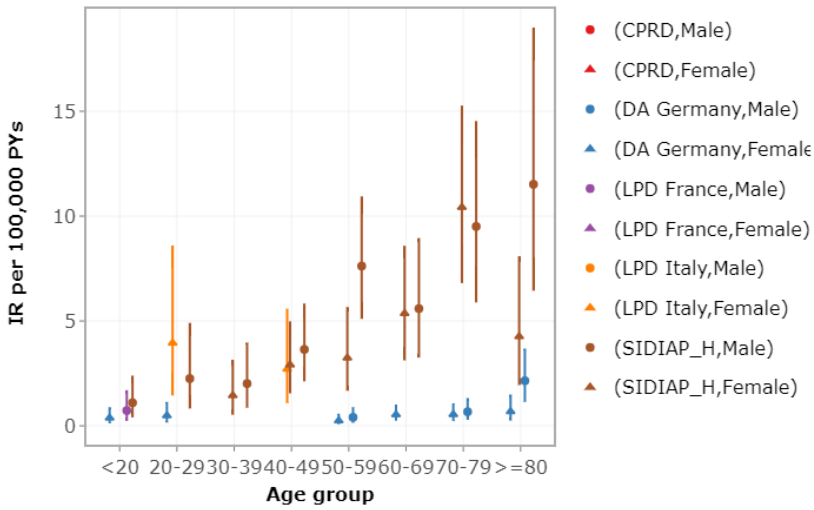
Heparin-induced thrombocytopenia was identifiable in 5 out of the 6 contributing databases, with background unadjusted rates ranging from 2.8 (2.6 to 3.1) in LPD FR to 7.8 (7.4 to 8.2)/100,000 py in DA Germany. As demonstrated in Figure 6, H-i-T was also more common in elderly people, and had higher rates in men than women.

Figure 6. Age and gender-specific background rates of heparin-induced thrombocytopenia



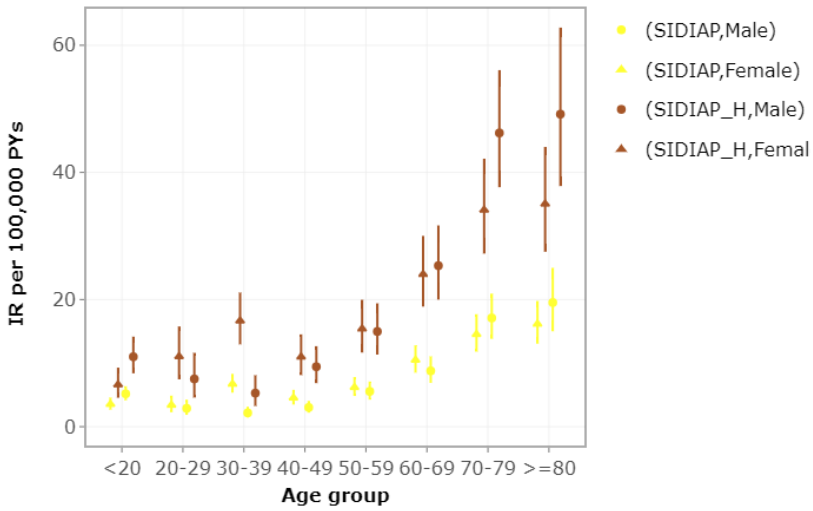
Disseminated intravascular coagulation was seen in 5 of the databases, but had a much higher unadjusted background rate in the only data source with linked inpatient data: rates ranged from 0.2 (0.1 to 0.3) in CPRD UK to 3.8 (3.3 to 4.3) in SIDIAP-H ES. Once again, rates increased with older age, and seemed higher in men than women in most of the reported analyses (see Figure 7).

Figure 7. Age and gender-specific background rates of disseminated intravascular coagulation



Finally, immune thrombocytopenia had an overall unadjusted background rate between 2.1 (1.8 to 2.5) in LPD FR to 16.7 (15.7 to 17.8) in SIDIAP-H ES. Again, hospital linkage seemed important for this outcome, with a much higher rate in SIDIAP-H than in the other data sources. This is illustrated, together with the observable age trends, in Figure 8, where data on the age-sex specific rates of immune thrombocytopenia in SIDIAP are compared to those observed in the subset of the database with linked hospital admission records (aka SIDIAP-H).

Figure 8. Age and gender-specific background rates of immune thrombocytopenia in SIDIAP and in the subset of the database with linked inpatient data (SIDIAP-H)

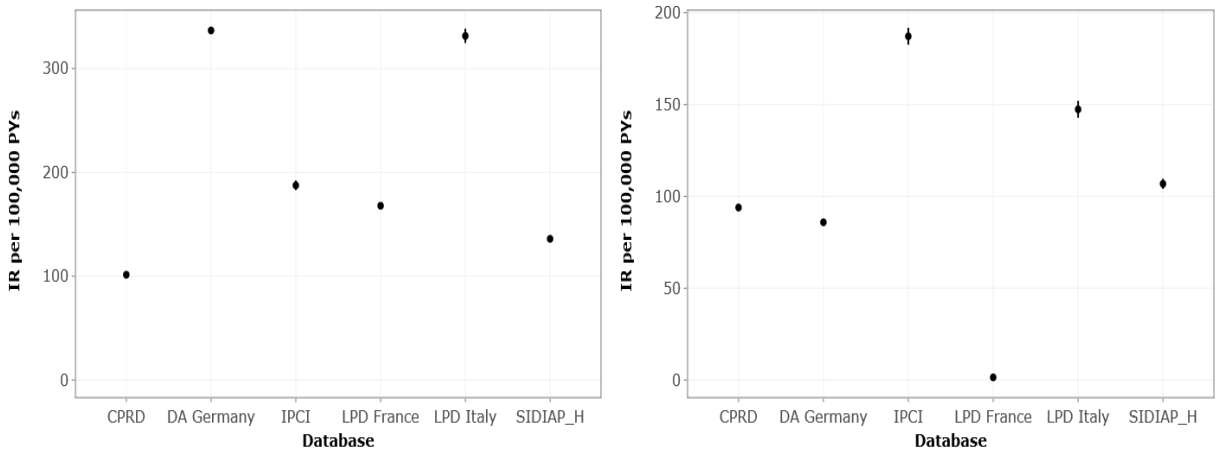


13.4 Database heterogeneity

13.4.1 Differences in coding practice across data sources

Substantial heterogeneity across databases was seen when using the narrow definition for DVT, which are not so obvious for the broad one (Figure 9). LPD France is an extreme example of this, as background incidence rates jumped from 1.4 for narrow to 167.8/100,000 for broad DVT.

Figure 9. Estimated incidence rates of DVT using a broad (left) vs narrow (right) definition



The most common source codes for our narrow DVT definition were "**Deep** vein **thrombosis**" (CPRD – Read code), "Phlebitis and **thrombo**phlebitis of other **deep** vessels of lower extremities" (DA Germany and LPD France – ICD-10 code), "Other venous embolism and **thrombosis** of inferior **vena cava**" (LPD Italy- ICD-9 code), and "Phlebitis and **thrombo**phlebitis of unspecified **deep** vessels of unspecified lower extremity" (SIDIAP – ICD-10-CM). All these include both key concepts of location (**deep** vein) and **thrombosis**.

Broadening the definition of DVT to codes that do not specify clearly the location ('deep') or thrombotic nature of the event had only modest effects on the results for some databases including CPRD (from 93.9 to 101.4/100,000) and IPCI (from 187.2 to 187.5/100,000). Conversely, this had a substantial impact for the following databases:

- In DA Germany, rates jumped more than 3-fold from 85.9 for narrow to 336.7/100,000 py. This change was driven by the most common code in the database (N=101,443 people, 178,408 records) being rather unspecific: ICD10 I80.3 "Phlebitis and thrombophlebitis of lower extremities, **unspecified**"

- In LPD France, rates jumped >100-fold, from 1.4 to 167.8/100,000 py. This increase was related to the inclusion of ICD10 I80.9 ("Phlebitis and thrombophlebitis of **unspecified** site") for DVT broad but not DVT narrow. This code alone accounted for 177,019 records (60,348 people), compared to only 4,178 (759 people) for the most commonly used DVT narrow code in this database: ICD10 I80.2 "Phlebitis and thrombophlebitis of other deep vessels of lower extremities".

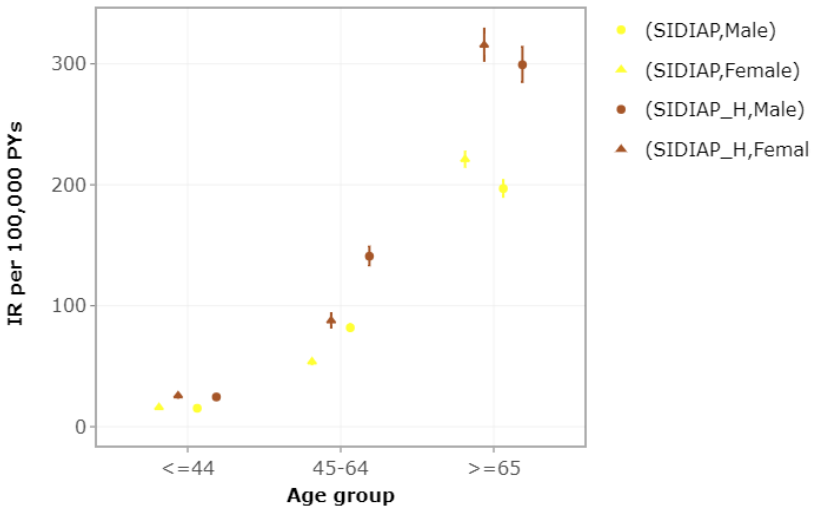
- In LPD Italy, rates changed from 147.4 to 331.5/100,000 py. This increase was driven by the inclusion of the most commonly recorded code for DVT broad in the database being again unspecific: ICD9CM 453.9 "Other venous embolism and thrombosis of unspecified site". This code alone appears a total of 147,654 times in the database, affecting 19,652 patients.

This provides one example of the underlying variation between the databases in how the presentation of a particular is captured in the medical record. While using a common data model facilitates the analysis of the data sources in a standardised manner, this underlying heterogeneity in the source data is still reflected to an important degree.

13.4.2 Impact of data linkage

Many of the events of interest can be expected to involve hospitalisation. Consequently, the degree to which diagnoses made in the hospital are reflected in the data sources used can be expected to have a large effect on the capture of events. We see this for SIDIAP vs SIDIAP-H: for our narrow definition, incidence rate of DVT increased from 64.5 in SIDIAP to 106.8/100,000 py in SIDIAP-H (Figure 10). Further, 44% of these DVT were in patients who were hospitalised at the time.

Figure 10. Estimated incidence rates for DVT (narrow definition) in SIDIAP and SIDIAP-H



13.5 TTS patient characterisation

The age and sex profiles of those with TTS are summarised in Table 4, and the prevalence of comorbidities and prior medication presented in Figure 11, along with those of the study populations.

Table 4. Characteristics of patients with thrombosis with thrombocytopenia

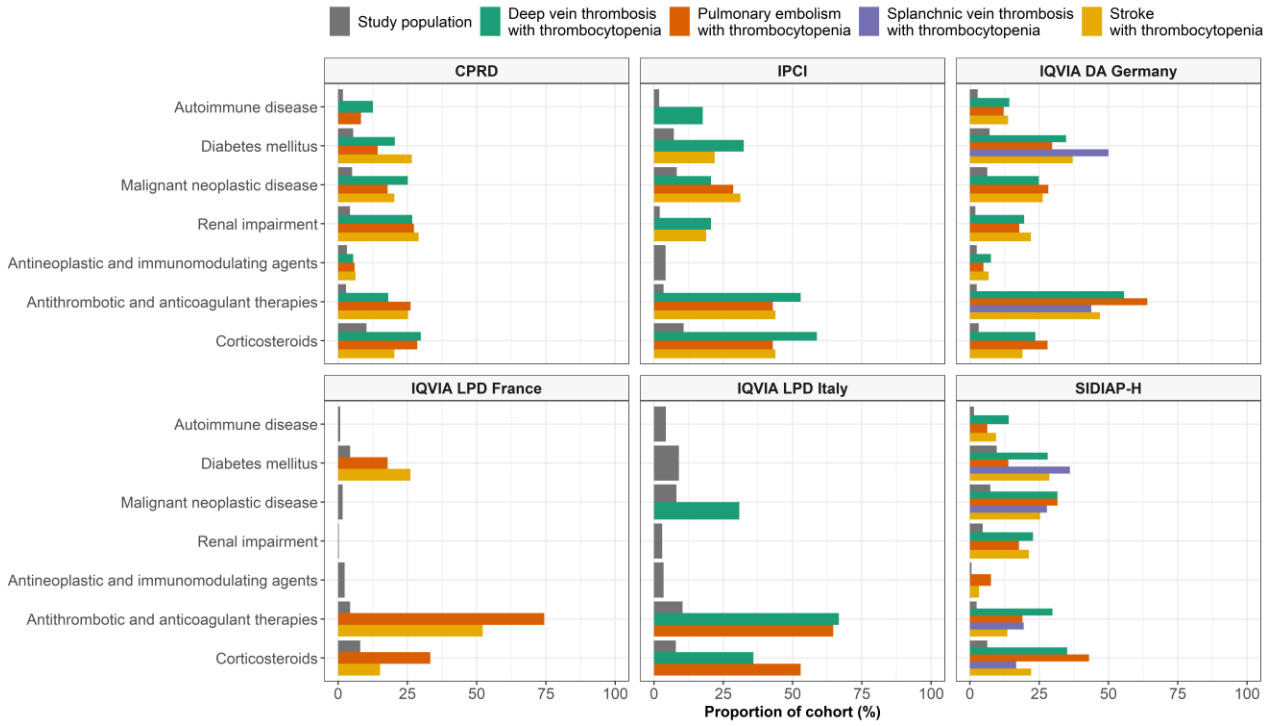
	N	**Age (Median [IQR])**	**Sex: Male (N [%])**
CPRD			
Study population	3,913,071	41 [22 to 59]	1,937,858 (49.5%)
Deep vein thrombosis with thrombocytopenia	127	70 [56 to 80]	70 (55.1%)
Pulmonary embolism with thrombocytopenia	84	71 [62 to 79]	50 (59.5%)
Splanchnic vein thrombosis with thrombocytopenia	5	59 [52 to 59]	<5
Stroke with thrombocytopenia	79	74 [66 to 81]	53 (67.1%)
IQVIA DA Germany			
Study population	8,459,098	52 [32 to 67]	3,589,506 (42.4%)
Deep vein thrombosis with thrombocytopenia	225	71 [60 to 80]	143 (63.6%)
Pulmonary embolism with thrombocytopenia	286	72 [62 to 80]	183 (64.0%)
Splanchnic vein thrombosis with thrombocytopenia	16	64 [60 to 73]	11 (68.8%)
Stroke with thrombocytopenia	369	76 [66 to 82]	244 (66.1%)
IQVIA LPD France			
Study population	3,951,633	48 [28 to 65]	1,669,415 (42.2%)

Pulmonary embolism with thrombocytopenia	39	72 [57 to 83]	23 (59.0%)
Stroke with thrombocytopenia	46	73 [66 to 82]	30 (65.2%)
IPCI			
Study population	1,299,288	44 [23 to 60]	636,386 (49.0%)
Deep vein thrombosis with thrombocytopenia	34	70 [54 to 81]	20 (58.8%)
Pulmonary embolism with thrombocytopenia	21	70 [54 to 73]	12 (57.1%)
Stroke with thrombocytopenia	32	74 [68 to 85]	22 (68.8%)
IQVIA LPD Italy			
Study population	1,066,230	52 [37 to 68]	426,758 (40.0%)
Deep vein thrombosis with thrombocytopenia	39	76 [62 to 82]	20 (51.3%)
Pulmonary embolism with thrombocytopenia	17	78 [69 to 81]	9 (52.9%)
SIDIAP-H			
Study population	1,909,814	43 [26 to 61]	945,873 (49.5%)
Cerebral venous sinus thrombosis with thrombocytopenia	6	68 [62 to 70]	<5
Deep vein thrombosis with thrombocytopenia	57	66 [57 to 77]	34 (59.6%)
Pulmonary embolism with thrombocytopenia	79	73 [62 to 78]	43 (54.4%)
Splanchnic vein thrombosis with thrombocytopenia	36	60 [52 to 67]	29 (80.6%)
Stroke with thrombocytopenia	244	72 [62 to 81]	156 (63.9%)

CPRD: Clinical Practice Research Datalink, IQVIA DA GERMANY: IQVIA Disease Analyser Germany, IQVIA LPD France: IQVIA Longitudinal Patient Data France, IPCI: Integrated Primary Care Information, IQVIA LPD Italy: IQVIA Longitudinal Patient Data Italy, SIDIAP-H: Information System for Research in Primary Care with hospital linkage

The median age of the six individuals with CVST with thrombocytopenia in SIDIAP-H was 68 years old. The median age of those with DVT with thrombocytopenia ranged from 66 to 76 across the databases, from 70 to 78 for PE with thrombocytopenia, from 59 to 64 for SVT with thrombocytopenia, and from 72 to 78 for stroke with thrombocytopenia. Men predominated in all the affected cohorts, accounting for 51.3% to 80.6% of those with different TTS in the contributing databases.

Figure 11. Comorbidities and medicines use in people with TTS



The prevalence of comorbidities and prior medication use was higher for patients with TTS than in the general population. In CPRD, for example, 1.8% of the source population had an autoimmune disease, 5.1% had a history of cancer, 5.5% had diabetes, 4.3% had renal impairment. These compared to 12.6%, 25.2%, 20.5%, and 26.8% for patients with DVT with thrombocytopenia. Similarly, while 2.9% of the study population were taking antithrombotic and anticoagulant therapies in the 6 months preceding their index date, 18.1% of patients with DVT with thrombocytopenia were.

14 Conclusions

We hereby report on the background incidence rates of thromboembolism, TTS, and coagulopathies of interest for the study of COVID-19 vaccine safety. The proportion of thromboembolic events with concomitant thrombocytopenia is well below 1/50 for DVT/PE, but higher (about 1/10 to 1/20) for CVT and visceral thrombosis. Although thrombocytopenia is common, HIT, DIC, and immune thrombocytopenia are relatively rare events, all with rates <20/100,000 overall. The rates of these increased with age and were higher in men than women in most of the performed analyses.

We found substantial heterogeneity in the background rates of the studied events in terms of age-sex and geography/database. Differences in coding practice explained a good part of the observed heterogeneity. The use of a common data model facilitates the identification of these differences, but does not resolve the underlying variation in coding practices across the network.

Linkage to external data sources (e.g. hospital records) was also an important contributor to the observed heterogeneity. While we can expect patient-level linkage of the other data sources to hospital level data to impact their results, it will not necessarily be to the same degree as in SIDIAP. In Catalonia, general practitioners are generally able to view patient hospital records during a consultation in a separate program to that used to view primary care records, with only the latter is captured in the SIDIAP database. Consequently, for patient care there is not a need to copy across what is seen in the system with the hospital record into the system with the primary care record which is captured by SIDIAP. This is not the case in other countries like the UK, where there is a need to add the information from letters received by general practitioners from secondary care facilities into the electronic medical record. Consequently, the impact of hospital linkage is likely to be context-dependent, as well as varying depending on the event under consideration.
Given the observed heterogeneity across databases, itt is recommended that a same population and data source should be used to analyse post-vaccine (observed) vs background (expected) rates for comparison.

The profile of patients with TTS after COVID-19 vaccination appears to differ to the typical profiles of those with TTS as seen in our data. While in this study we have seen those with TTS to typically be older than the general population, more commonly male, and with more comorbidities and greater prior medication use, initial studies describing the profiles of patients with vaccine-

induced TTS have most often presented the cases of people who were aged under 60, more often female, and with relatively few comorbidities described. While these case series of TTS after vaccination are small and their profiles may reflect the particular characteristics of those who were first to receive a vaccine against SARS-CoV-2, this dissimilarity in patient profiles of those with TTS in previous years and those for whom it has been reported following a vaccination is notable. These differences highlight the need for further research into the nature of post-vaccination TTS before background rates of TTS can be used as a benchmark for post-marketing safety surveillance.

15 References

1 Bikdeli B, Madhavan M V, Jimenez D, *et al.* COVID-19 and Thrombotic or Thromboembolic Disease: Implications for Prevention, Antithrombotic Therapy, and Follow-Up: JACC State-of-the-Art Review. *J Am Coll Cardiol* 2020;**75**:2950–73. doi:https://doi.org/10.1016/j.jacc.2020.04.031

2 Klok FA, Kruip MJHA, van der Meer NJM, *et al.* Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thromb Res* 2020;**191**:145–7. doi:10.1016/j.thromres.2020.04.013

3 Lodigiani C, Iapichino G, Carenzo L, *et al.* Venous and arterial thromboembolic complications in COVID-19 patients admitted to an academic hospital in Milan, Italy. *Thromb Res* 2020;**191**:9–14. doi:10.1016/j.thromres.2020.04.024

4 Llitjos J-F, Leclerc M, Chochois C, *et al.* High incidence of venous thromboembolic events in anticoagulated severe COVID-19 patients. *J Thromb Haemost* 2020;**18**:1743–6. doi:10.1111/jth.14869

5 Hessami A, Shamshirian A, Heydari K, *et al.* Cardiovascular diseases burden in COVID-19: Systematic review and meta-analysis. *Am J Emerg Med* Published Online First: 18 January 2021. doi:10.1016/j.ajem.2020.10.022

6 Burn E, Tebe C, Fernandez-Bertolin S, *et al.* The natural history of symptomatic COVID-19 in Catalonia, Spain: a multi-state model including 109,367 outpatient diagnoses, 18,019 hospitalisations, and 5,585 COVID-19 deaths among 5,627,520 people. *medRxiv* 2020;:2020.07.13.20152454. doi:10.1101/2020.07.13.20152454

7 Petrilli CM, Jones SA, Yang J, *et al.* Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* 2020;**369**. doi:10.1136/bmj.m1966

8 Reilev M, Kristensen KB, Pottegaard A, *et al.* Characteristics and predictors of hospitalization and death in the first 11 122 cases with a positive RT-PCR test for SARS-CoV-2 in Denmark: a nationwide cohort. *Int J Epidemiol* Published Online First: 2020. doi:10.1093/ije/dyaa140

9 Zhou F, Yu T, Du R, *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;**395**:1054–62. doi:10.1016/S0140-6736(20)30566-3

10 Docherty AB, Harrison EM, Green CA, *et al.* Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020;**369**. doi:10.1136/bmj.m1985

11 Gupta S, Hayek SS, Wang W, *et al.* Factors Associated With Death in Critically Ill Patients

With Coronavirus Disease 2019 in the US. *JAMA Intern Med* Published Online First: 15 July 2020. doi:10.1001/jamainternmed.2020.3596

12 Williamson EJ, Walker AJ, Bhaskaran K, *et al.* OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature* Published Online First: 2020. doi:10.1038/s41586-020-2521-4

13 Lowe GDO. Common risk factors for both arterial and venous thrombosis. *Br J Haematol* 2008;**140**:488–95. doi:https://doi.org/10.1111/j.1365-2141.2007.06973.x

14 Previtali E, Bucciarelli P, Passamonti SM, *et al.* Risk factors for venous and arterial thrombosis. *Blood Transfus* 2011;**9**:120–38. doi:10.2450/2010.0066-10

15 Collard D, Nurmohamed NS, Kaiser Y, *et al.* Cardiovascular risk factors are independently associated with COVID-19 mortality: a prospective cohort study. *medRxiv* 2020;:2020.10.01.20205229. doi:10.1101/2020.10.01.20205229

16 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020;**369**:m1328. doi:10.1136/bmj.m1328

17 Reps JM, Schuemie MJ, Suchard MA, *et al.* Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *J Am Med Informatics Assoc* 2018;**25**:969–75. doi:10.1093/jamia/ocy032

18 Cheng MP, Papenburg J, Desjardins M, *et al.* Diagnostic Testing for Severe Acute Respiratory Syndrome–Related Coronavirus 2. *Ann Intern Med* 2020;**172**:726–34. doi:10.7326/M20-1301

19 Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics : Competing risks and multi-state models. 2007;:2389–430. doi:10.1002/sim

20 Leek JT, Peng RD. What is the question? *Science (80-)* 2015;**347**:1314–5. doi:10.1126/science.aaa6146

21 Kahn MG, Callahan TJ, Barnard J, *et al.* A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Washington, DC)* 2016;**4**:1244. doi:10.13063/2327-9214.1244