NATURAL HISTORY OF COAGULOPATHY AND USE OF ANTI-THROMBOTIC AGENTS IN COVID-19 PATIENTS AND PERSONS VACCINATED AGAINST SARS-COV-2 – REPORT 3: INCIDENCE RATES OF ARTERIAL THROMBOEMBOLIC EVENTS (ATE) AND VENOUS THROMBOEMBOLIC EVENTS (VTE) IN COVID-19 PATIENTS, THEIR IMPACT ON PROGNOSIS, AND THE PREDICTION OF THESE EVENTS

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

Document Status: Version 2.0 -

Date: 8th October 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	1.1
Date of last version of protocol	07/06/2021
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 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 report contains results for Objectives 4 to 11.**
Country(-ies) of study	Italy, France, Germany, Netherlands, Spain, and United Kingdom.
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

LIST OF FIGURES

LIST OF TABLES

2 List of abbreviations

Abbreviation	Name
AIC	Akkaike Information Criterion
ATC	Anatomical Therapeutic Chemical Classification
ATE	Arterial thromboembolic events
AUROC	Area Under the Receiver Operating Characteristic curve
BIC	Bayesian Information Criterion
CDM	Common Data Model
COVID-19	Coronavirus disease-2019
CPRD	Clinical Practice Research Datalink
DA	Disease Analyzer
DE	Germany
EHR	Electronic Health Record
ES	Spain
FR	France
IPCI	Integrated Primary Care Information
IT	Italy
LPD	Longitudinal Patient Data
MACE	Major cardiovascular events
NL	Netherlands
OMOP	Observational Medical Outcomes Partnership
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SIDIAP CMBD-AH	The Information System for Research in Primary Care (SIDIAP) fully linked to the regional (Catalonia) hospital admissions data generated by the "programme of data analyses for research and innovation in health" (PADRIS for its acronym in Catalan language)
UK	The United Kingdom of England, Northern Ireland, Scotland, and Wales
VTE	Venous thrombo-embolism

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 and Data Scientists

Dr Edward Burn (University of Oxford and IDIAPJGol)

Mr Henrik John (EMC)

Data controllers/analysts

Marcel de Wilde (EMC)

Mees Mosseveld (EMC)

Kristin Kostka (Roux Institute, North-Eastern University, formerly IQVIA)

Database Stewards

Prof Johan van der Lei (IPCI)

Dr. Christian Reich (IQVIA)

Dr. Sarah Seager (IQVIA)

Dr. Talita Duarte-Salles (SIDIAP, HM)

Dr. Antonella Delmestri (CPRD GOLD and AURUM)

4 Abstract

Version and Date: Version 2.0, 11th September 2021

Name and affiliation of main authors:

Daniel Prieto-Alhambra (University of Oxford, Erasmus MC)

Edward Burn (University of Oxford, Idiap Jordi Gol)

Rationale and background:

Risks of venous thromboembolic events (VTE) and arterial thromboembolic events (ATE) have been reported to be increased in the days following a diagnosis of COVID-19. The incidence of these events is not yet well established, nor is their impact on patient prognosis, It is not well known whether classic risk factors for these events are also relevant for persons with COVID-19. An improved understanding of this could help target patients for preventative therapies including antithrombotic or anticoagulants.

Research question and objectives

The current report contains results on eight objectives as part of a larger study. The objectives covered here are Objectives 4 to 11:

- *4) To estimate the incidence of VTE in patients with COVID-19 at 30-, 60-, and 90-days.*
- *5) To calculate the risks of COVID-19 worsening stratified by the occurrence of VTE.*
- *6) To assess the impact of risk factors on the rates of VTE among patients with COVID-19.*
- *7) To develop and externally validate prediction models for VTE in COVID-19 patients.*
- *8) To estimate the incidence of ATE in patients with COVID-19 at 30-, 60-, and 90-days.*
- *9)To calculate the risks of COVID-19 worsening stratified by the occurrence of ATE.*
- *10) To assess the impact of risk factors on the rates of ATE among patients with COVID-19.*
- *11) To develop and externally validate prediction models for ATE in COVID-19 patients.*

Study design

We performed a European international network cohort study using data from 5 European countries: France, Germany, Netherlands, Spain, and the United Kingdom (UK).

Population

People with a specific clinical diagnosis of COVID-19 or a positive PCR test against SARS-CoV-2 were included. People with <1 year of data visibility before index date were excluded. Further,

prediction algorithms related to Objectives 7 and 11 excluded previously vaccinated people to increase their current and future clinical usefulness in Europe.

Variables

Primary outcomes of interest included: deep vein thrombosis (DVT), pulmonary embolism (PE), and the composite of both (VTE); myocardial infarction (MI), ischaemic stroke IS, and the composite of both (arterial thromboembolism [ATE]). Demographics, medical history, and medication/s use were extracted for the analysis of risk factors and for prediction modelling purposes.

Data sources

Primary care records from Netherlands (IPCI), Spain (SIDIAP), and the UK (CPRD AURUM); ambulatory data from France (IQVIA LPD France) and Germany (IQVIA DA Germany). SIDIAP (primary care) was further linked to inpatient data (CMBD-AH) from Catalonia (from here on SIDIAP CMBD-AH) to maximise completeness. All contributing data sources were mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model. IQVIA LPD Italy was excluded from Objectives 4-11 due to the limited number of COVID-19 patients identifiable in this dataset (n<5,000) and their limitation of capturing data before October 2020.

Study size

The following number/s of participants were included: 43,151 (IPCI NL), 484,810 (SIDIAP CMBD-AH ES), 439,958 (CPRD UK), 45,370 (IQVIA DA DE), and 103,086 (IQVIA LPD FR).

Data analyses

Cumulative incidence of VTE (Objective 4) and ATE (Objective 7) were estimated at 30, 60, and 90-day windows after COVID-19 diagnosis or positive PCR test, overall and stratified by age-sex and by period (before vs after September 2020). Kaplan-Meier and, to account for competing risk of death, Cumulative Incidence Functions were estimated. The association between the occurrence of VTE/ATE and worsening (defined by hospitalisation or death) related to Objectives 5 and 8 was modelled using multi-state survival models in SIDIAP CMBD-AH. Overall and cause-specific hazard ratios (HR) were estimated for the association between pre-specified determinants and VTE/ATE risk, unadjusted and after age-sex adjustment. Finally, prediction models were trained using CPRD UK and externally validated in all other databases. All analyses were conducted as a distributed network.

Results

All study results for Objectives 4, 6, 7, and 9, including characterisation, VTE/ATE cumulative incidences and hazard ratios, are reported in full here: https://livedataoxford.shinyapps.io/ThromboembolismInCovid19_working/

This is the largest study on this topic to date, including a total of 1,116,375 participants. Mean age ranged from 43 (ES and UK) to 47 (NL and DE), and % of men from 43.7% (FR) to 47.1% (ES). The cumulative incidence of VTE on 30-, 60-, and 90 days ranged from 0.08%, 0.13%, 0.16% (FR) to 0.54%, 0.62%, 0.66% (ES); similarly, figures for ATE ranged from 0.02%, 0.03%, 0.04% (FR) to 0.48%, 0.56%, 0.61% (ES). Age-sex specific cumulative incidences were higher for the older vs younger, and for men vs women in most analyses.

In the multistate model informed by data from ES, an outpatient diagnosis of VTE after COVID-19 was associated with an increased risk of both hospitalisation (adjusted HR: 1.36 [95% CI: 0.95 to 1.96]) and even more so with death without a COVID-19 hospitalisation (4.42 [3.07 to 6.36]). VTE on or after admission with COVID-19 was also associated with an increased risk of death (1.63 [1.39 to 1.90]). ATE prior to a COVID-19 hospitalisation was not associated with COVID-19 hospitalisation (1.05 [0.89 to 1.25]), but it related strongly to an increased risk of death without hospitalisation (3.16 [2.65 to 3.75]). ATE on or after admission with COVID-19 was associated with an increased mortality (1.93 [1.57 to 2.37]).

Cox models showed older age was associated with VTE and ATE risks in a non-linear fashion, with small linear effects in ages <60 but a stronger effect size in older age. Male sex was associated with an increased risk of VTE and ATE in most databases. A history of renal impairment, cancer, diabetes, and previous use of systemic glucocorticoids, or of antithrombotics were associated with higher risk of VTE in most databases despite age-sex adjustment. Similarly, smoking, renal impairment, obesity, hypertension, heart disease, diabetes mellitus, and the previous use of systemic glucocorticoids were associated with ATE risk in age-sex adjusted models.

An algorithm based on age and sex was able to predict VTE and ATE with good discrimination (AUC>0.8) and calibration in patients aged <65 in CPRD UK. The resulting algorithms were externally validated with similar discrimination in DE, FR, NL, and ES. The addition of other risk factors did not improve performance in most scenarios.

Conclusions

The cumulative incidence of VTE and ATE post-COVID varied across European databases, probably due to previously described heterogeneity in source coding, healthcare settings included,

and linkage availability. VTE and ATE were associated with COVID-19 worsening as defined by hospitalization and mortality.

Different pre-specified risk factors appeared associated with the study outcomes in age-sex adjusted models. For example, asthma was associated with VTE but not ATE. Conversely, obesity was associated with ATE but less consistently with VTE. Confounding by indication and channelling bias preclude the causal interpretation of the observed associations, particularly for the study of prevalent therapies. Further research including randomised controlled trials are needed to study the efficacy of antithrombotic and anticoagulant therapies to prevent VTE and ATE post COVID-19.

Finally, an equation based on age and sex predicted VTE and ATE accurately in COVID-19 patients aged <65, but not in older ones. The addition of more risk factors did not improve the discrimination of the resulting algorithm. More work is needed to accurately identify elderly COVID-19 patients at highest risk of VTE or ATE events for trial recruitment and/or targeted anticoagulant or antiaggregant treatment/s.

5 Amendments and updates

Number	Date	Amendment or update	Reason
1 (Ver 1.1)	*April 2021*	*Addition of vaccinated cohorts to the study population* *Incidence rate ratios to compare post-covid and post-vaccine to background (historical) rates*	*Safety signals of some COVID vaccine/s associated with excess risk of coagulopathy / thromboembolism*
2 (Ver 1.2)	*7th June 2021*	*Standardization added to incidence rate ratios*	*Differences noted between vaccinated vs background cohort*

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 The occurrence of venous and arterial thromboembolic events in patients with COVID-19

Coronavirus disease-2019 (COVID-19) may result in thrombotic disease, both in the venous and arterial circulations, due to excessive inflammation, platelet activation, endothelial dysfunction, and stasis.[1] Indeed, a number of studies have already reported high rates of arterial and venous thromboembolic events among patients hospitalised with COVID-19. Case series of patients admitted to an intensive care unit (ICU) have, for example, described the high incidence of such events. In a case series of COVID-19 patients admitted to ICU in the Netherlands, the incidence of thrombotic complications was found to be 31%,[2] while a similar case series from a hospital in Italy found the incidence of thromboembolic events to be 28%.[3] Meanwhile, the rate of venous thromboembolism was found to be as high as 69% for a case series from two French intensive care units (ICU).[4]

Previous studies assessing the incidence of thromboembolic events in COVID-19 have, however, typically been based on relatively small study populations and have predominantly focused only on hospitalised or even ICU-admitted patients, with little evidence available on the wider population of COVID-19 patients. Consequently, uncertainty remains around the incidence of thromboembolic events among patients with COVID-19. Study cohorts derived from routinely-collected healthcare data can provide the requisite breadth of data capture and sample size to address this research gap. Moreover, such data can be used to consider the incidence of thromboembolic events for particular groups of interest, for example those with a history of thromboembolic events.

7.2 Thromboembolic events and worsening in patients with COVID-19

COVID-19 patients with a thromboembolic events appear to be at increased risk of worse outcomes, with a recent systematic review finding a strong association between cardiovascular and thromboembolic events and poor prognosis in COVID-19.[5] As with the incidence of the events themselves, routinely-collected data can be used to describe the risk of worsening in COVID-19 among those with and without a thromboembolic event.

7.3 Risk factors for thromboembolic events in patients with COVID-19

Various patient factors have been associated with worse outcomes in COVID-19. Older age, male sex, hypertension, diabetes, and overweight/obesity have all been associated with an increased risk of hospitalisation and mortality in COVID-19.[6–12] Many of these same features have also previously been shown to predispose individuals to thromboembolic events.[13,14] Indeed in one study a set of pre-existing cardiovascular risk factors were associated with mortality in COVID-19,

independent of patients´ age and sex.[15] The associations between such risk factors and thromboembolic events among patients with COVID-19 has though yet to be elucidated in detail.

7.4 Predicting thromboembolic events in patients with COVID-19

Prediction models that combine various patient features can be used to estimate individuals´ personalised risks of adverse outcomes in COVID-19. A prediction model of thromboembolic events among patients with COVID-19 is lacking. If available and with good performance, such models would be a valuable tool in the management of COVID-19. Numerous prediction models have been developed for COVID-19, but many have been limited by small sample sizes, a lack of representative study populations, and an absence of external validation.[16] Using routinely-collected data can offer a solution to these potential issues, especially when using data mapped to a common data model which would allow for the development and external validation of models to be done in both a timely and reproducible manner.[17]

8 Research question and objectives

The overall study objectives are:

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 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.

Whilst Objective 1-3 were reported in previous documents, we here report on the results from Objectives 4-11.

9 Research methods

9.1 Study design

Network cohort study using routinely collected health care (aka "real world") 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 were used to inform the analyses (see section 9.4 Data Sources below for more details).

9.2.2 Study period

The study period differed for the included databases, and covered from March/2020 (all data sources) to end of January 2021 (UK), March 2021 (FR, NL), April 2021 (DE), and end of June 2021 (ES). Italian data was not updated beyond September 2020 and were therefore not used for this report. After exploring Hospitales Madrid (HM), we decided not to include this database as it had no granularity on diagnosis dates: all diagnoses whilst inpatient were recorded as occurring on the day of admission, limiting any modelling of such data for this specific study. All analyses were stratified as before vs from September 2020.

9.2.3 Study cohorts

The following study cohorts were defined:

1. Primary cohorts
    - *COVID-19 PCR+ or diagnosis*: defined by either a clinical diagnosis of COVID-19 ("narrow definition") or a positive PCR test. The first of either of these is used as index date.
    - *Hospitalised with a COVID-19 PCR+ or diagnosis:* SIDIAP CMBD-AH had linkage to hospital admissions data. We took advantage of this to define a 'hospitalised' cohort defined by the criteria above combined with a hospital admission. For this cohort, a positive PCR or COVID-19 diagnosis is required to be observed between 21 days prior to admission up to 3 days after admission. Index date is the date of hospitalization.
2. Additional cohorts: Given the heterogeneity in testing and diagnostic practices, we defined the following additional cohorts based on the available information in each of the contributing databases:
    - *COVID-19 diagnosis broad*: defined by a broad algorithm/list of codes for diagnosis

- *COVID-19 diagnosis narrow:* defined by clinical diagnosis as in the primary cohort
- *COVID-19 PCR positive test:* defined by a positive PCR test as in primary cohorts
- *COVID-19 positive test:* defined by a positive test, regardless of type of test i.e. including lateral flow (rapid antigen) tests, antibody tests, or PCR
- *Hospitalised with a COVID-19 PCR positive test:* defined by a hospital admission with a positive PCR test observed between 21 days prior to and 3 days after date of admission.

All the above cohorts were generated with a 1-year run-in in the database, with this criterion removed in additional (sensitivity) analyses. This report includes results from primary cohorts with 1-year run-ins, with all additional cohorts and sensitivity analyses reported in Appendix 1.
The developed (and validated) prediction algorithms (Objectives 7, 11) were based on the primary cohort (COVID-19 PCR+ or diagnosis) who were unvaccinated on index date, and who were included from September 2020. This was done to maximise the clinical usefulness of the developed tools in the current circumstances.

9.2.4 Follow-up

Follow-up went from cohort-specific index date and until the first of: outcome of interest, death, loss to follow-up, 90-day post-index, or end of the study period. Study outcomes are reported using 30-, 60-, and 90-day time windows post-index, with higher granularity (daily timepoints) reported in an accompanying Shiny app.

9.3 Variables

9.3.1 Study outcomes

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

9.3.1.1 Venous thromboembolic events

Venous thromboembolic events (VTE) were identified by diagnostic codes for pulmonary embolism or deep vein thrombosis. Additionally, pulmonary embolism (PE) and deep vein thrombosis (DVT) were assessed and reported separately as independent events.

9.3.1.2 Arterial thromboembolic events

Arterial thromboembolic events (ATE) were identified as a composite of an acute myocardial infarction or acute ischemic stroke. In addition, acute myocardial infarction (MI) and acute ischemic stroke (IS) were assessed separately.

9.3.1.3 Additional events

We identified the occurrence of the following additional outcomes based on current clinical knowledge of cardiovascular complications in COVID-19 patients:

1. Cardiac arrhythmia
2. Angina/chest pain
3. Heart failure
4. Haemorrhagic stroke
5. Stroke (regardless of aetiology)
6. Major cardiovascular events (MACE), including any of the following: heart failure, acute myocardial infarction, stroke, or the occurrence of sudden cardiac death.
7. Sudden cardiac death
8. Ventricular arrhythmia / cardiac arrest
9. Death

We analysed these additional events and report on each of them in Appendix 1. Death was also used for competing risk modelling where reliable and complete data on mortality were available (ES, NL, and the UK).

9.3.2 Characteristics of study participants

9.3.2.1 Demographics

Age at index date and sex were characterized.

9.3.2.2 Health conditions and medications use pre-index date

Prevalence of key pre-specified health conditions as recorded any time before index date are reported for the primary cohort populations. Medications use as recorded in the six months (183 days prior to 4 days prior) before index date, in alignment with an ongoing similar study by FDA-Sentinel, are also reported.

The list of selected conditions and medicines has been previously reported in previous documents, and includes:

• Autoimmune disease
• Antiphospholipid syndrome
• Thrombophilia

• Asthma
• Atrial fibrillation
• Malignant neoplastic disease
• Diabetes mellitus
• Obesity
• Heart disease
• Hypertensive disorder
• Renal impairment
• COPD
• Dementia
• *Any of the above*

The list of selected medicines used in the 6-month to 4 days before index date 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
• Immunoglobulins
• *Any of the above*

In addition to conditions and medicines, smoking status was assessed based on diagnostic codes and measurements recorded any time before index date indicating a person to have been a smoker.

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. Based on results in previous reports, we applied for full linkage of SIDIAP to the regional hospital admissions database (CMBD-AH for its acronym in Catalan language), which were obtained in time for this analysis. We have renamed the data source as SIDIAP CMBD-AH to reflect the full linkage of both databases for this study.

In turn, LPD IT (primary care records from Italy) has not been updated beyond September 2020, and includes a very small sample size. Although useful for previously reported background (historical) rate analyses, we excluded this database from the current report due to these limitations.

Table A: Data sources accessible for analysis

Database	**Managing Organization**	**Country**	**Description**
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 AURUM	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 AURUM contains data contributed by GP practices using EMIS® 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 CMBD-AH	IDIAP Jordi Gol	Spain	SIDIAP CMBD-AH is the results of linking the entire SIDIAP database to regional (Catalan) administrative data related to hospital admissions equivalent to the Conjunt Minim Basic de Dades a l'Alta Hospitalaria (CMBD-AH) for all hospitals in the Catalan healthcare system. This linkage is available on a study-per-study basis, and was provided by the PADRIS programme for this specific analysis, covering up and until 30th June 2021. Both SIDIAP and the linked PADRIS data were linked and jointly mapped to the OMOP CDM for further processing.

9.5 Study size

For each database, all individuals that satisfied the eligibility criteria for any of the listed study cohorts were included. The number of people eligible in the primary cohorts 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

All analytical code for this study was written in R. Each data partner executed the same 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 contains the code used for Objectives 4, 5, 6, 8, 9, and 10
- Code to train prediction models are available at
    - Baseline: https://github.com/mi-erasmusmc/CovCoagBasePrediction
    - Parsimonious: https://github.com/mi-erasmusmc/CovCoagEmaPrediction
- Validation packages are available at
    - Baseline: https://github.com/mi-erasmusmc/CovCoagBaseValidation
    - Parsimonious: https://github.com/mi-erasmusmc/CovCoagEmaValidation

9.7.2 Descriptive statistics

The observed characteristics of primary cohorts are reported in the current report. Additional cohorts, including the subsets of those affected by each of the study outcomes, are characterised in our interactive web application, under the 'Patient profile' tab:
https://livedataoxford.shinyapps.io/Covid19Thrombosis/

9.7.3 Cumulative incidence (Objectives 4, 8)

We estimated the cumulative incidence for each of the study outcomes in each of the study cohorts, as well as stratified by age, sex, and period (before vs from September 2020), in line with the ongoing FDA Sentinel-based study. Cumulative incidence of events of interest was estimated using the Kaplan-Meier approach and, for those databases where mortality was well captured, and from the cumulative incidence function (CIF) taking the competing risk of mortality into account. Estimates of cumulative incidence, with 95% confidence intervals, are provide at days 30-, 60-, and 90 after index date. The main analyses of ES, NL and UK data used cumulative incidence functions (CIF), whereas the Kaplan-Meier (KM) approach was used for DE and FR(where mortality data is unreliable).

The results reported in this document are based on CIF (ES/NL/UK) or KM (DE/FR) of the primary study outcomes (VTE, ATE, and the individual events included) based on the primary cohorts with 1+ years of data visibility. Analyses of additional outcomes, with/without 1-year washout, using KM instead of CIF (for ES/NL/UK), and using more granular (daily) time periods are reported in the accompanying interactive web application, under 'Cumulative incidence', and in Appendix 1.

9.7.4 Association between VTE/ATE and worsening (Objectives 5, 9)

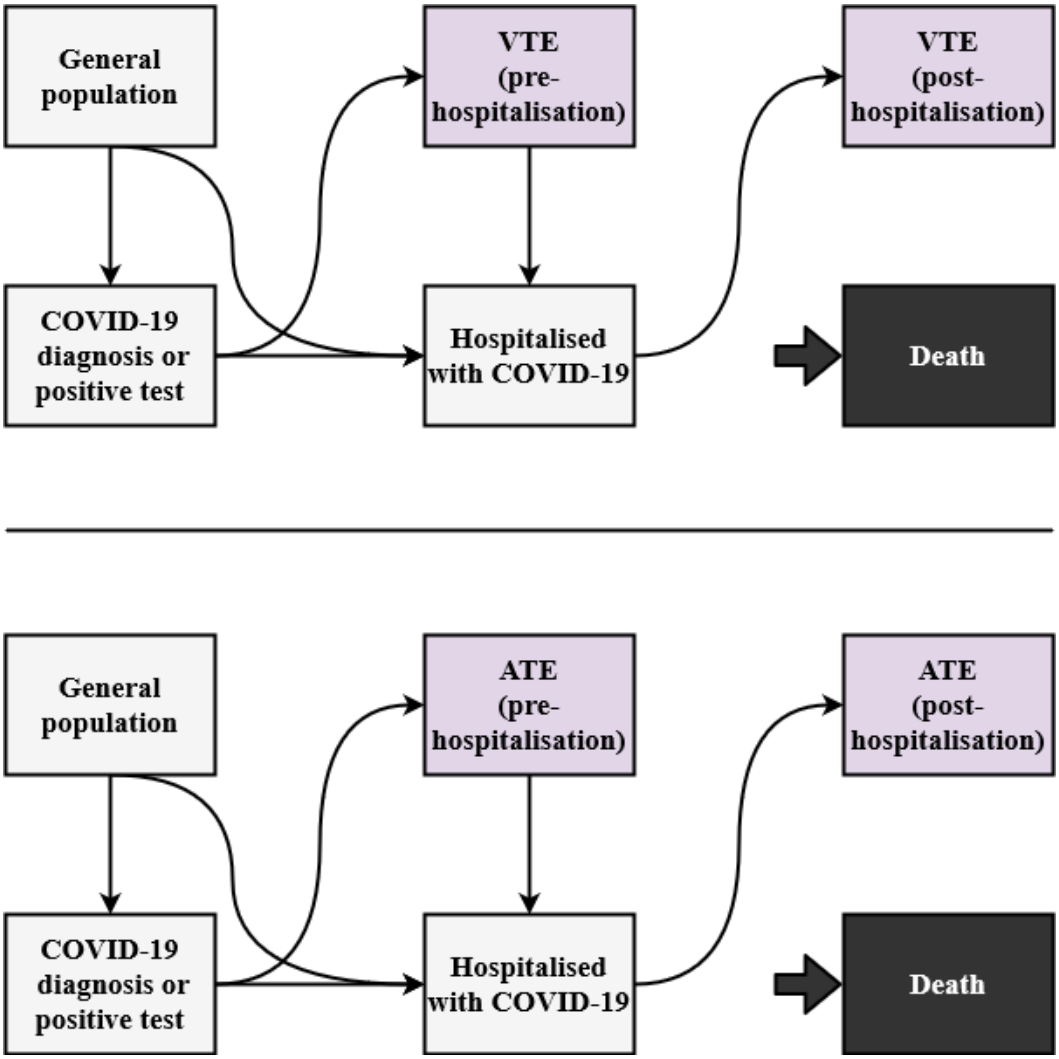
Multi-state = models provided the framework for describing patient trajectories. These models were informed by data from ES, with SIDIAP CMBD-AH the sole dataset that captured the full general population, outpatient diagnoses of COVID-19 and testing for SARS-CoV-2, and linked hospitalisation and mortality data. The model outline is shown below in **Figure 1**.

For this model, the full SIDIAP CMBD-AH dataset was used to identify the starting population of persons in the general population. This study population was defined as all persons in SIDIAP as of the 1st September 2020 with: 1) no prior history of COVID-19, 2) no record of VTE or ATE in the year prior, and 3) at least one year of prior observation time available. This starting population beginning in the "General Population" state could then transition through the model as time progressed, with the model running from the 1st September 2020 up until 30th June 2021. Follow-up time was censored if a person had a VTE or ATE prior to a diagnosis of COVID-19 or positive PCR test, but otherwise ran up to whichever came first of death (the final, absorbing state in the model), exit from the database, or the end of the study period.

The characteristics of the starting study population and the cohorts of people making each of the tranistions in the model are summarised. The cumulative incidence for each of these transitions is described based on the Cumulative Incidence Functions, with alternative transitions from a given state considered as a competing risk for the event of interest. Cumulative incidence of transitions out of the general population are described for the full study period, wheras we describe cumulative incidence at 90-days for all other states.

To assess the impact of ATE and VTE, pre- or post-hosptialisation, these events were considered as time dependent exposures. That is the two ATE and VTE states are removed from the model, with these occurrences then considered as explanatory factors for transitions rather than states in themselves. In this manner, the association between VTE ocurring prior to any COVID-19 hospitalisation and the risk of hospitalisation with COVID-19 and death (without any prior COVID-19 hospitalisation) was evaluated using Cox models with VTE prior to any COVID-19 hospitalisation included as an the sole explanatory factor (unadjusted model) or along with age and sex (adjusted model). Similarly the association between VTE at the time of or after hospital admission and death was assessed. The impact of ATE on patient trajectories was studied in an analogous manner.

Figure 1 - Framework for multi-state modelling of the association between VTE/ATE occurrence and COVID-19 worsening



9.7.5 Association between pre-specified risk factors and VTE/ATE (Objectives 6, 10)

We analysed each of the pre-specified risk factors above (Section 9.3.2.2) individually with each of the study outcomes and in each individual database. We also analysed the association between sex and each of the study outcomes, unadjusted and age-adjusted. The association between age and each of the outcomes was studied using either a linear term of with non-linearity incorporated using restricted cubic splines, with AIC used to identify the best fit to the observed data.

For each of these analyses, we used cause-specific proportional-hazards Cox models to estimate Hazard Ratios (and 95% Confidence Intervals) for each study outcome according to risk factor status. Unadjusted and age-sex adjusted models are also reported.

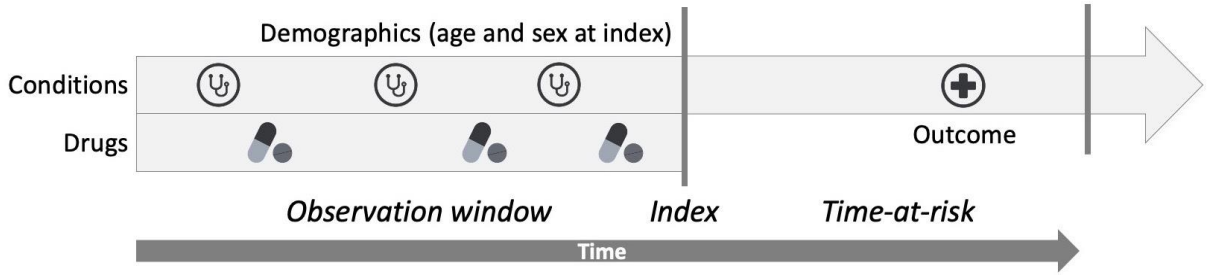
9.7.6 Development and validation of algorithms for the prediction of VTE/ATE in COVID-19 patients (Objectives 7, 11)

We used the patient-level prediction framework for model development and validation [JM Reps et al. JAMIA 2018]. This framework enables the development of analysis packages in R that can be shared across data sites mapped to the OMOP CDM. **Figure 2** illustrates the modeling approach used to predict each of the outcomes given a set of predictors from the condition and drug domain, recorded in an observation window prior to the index date. Three models were planned per protocol:

- Baseline models, including only age and sex
- Parsimonious models: based on age, sex and pre-specified risk factors (Section 9.3.2.2)
- Data-driven models, including all potential comorbidities and medicines use available in the utilized databases, and selected using regularization

The latter did not improve the performance of parsimonious models despite increased complexity in internal validation exercises, and were therefore not pursued further.

Figure 2 - Patient-level prediction of outcome in a time window following the index date using predictors from the condition and drug domain



We used a train-test split by person to perform internal validation. Each person appeared only once in the datasets, because we only use their earliest index date. In the development cohort, a random sample of 75% of patients was used to develop the prediction models and the remaining 25% of patients were used to internally validate the models. We trained models using LASSO regularized logistic regression with 3-fold cross validation to learn the optimal regularization hyperparameter through an adaptive search.

To evaluate the performance, we estimated the discrimination of the model using the area under the receiver operating characteristic curve (AUROC) and the model calibration (Eavg and age-sex stratified observed vs expected risk plot). The AUROC indicates the probability that for two

randomly selected patients, the patient who gets the outcome will be assigned a higher risk. The model calibration is presented in a plot to examine agreement between predicted and observed risk across deciles (or centiles) of predicted risk. On top of graphically, calibration was assessed using Eavg, a single value metric, which uses the average absolute difference between observed and predicted probabilities (where smaller values are better) [Austin PC et al. Stats in Med 2019].

For each dataset, models were developed in a training set based on CPRD AURUM data, and evaluated in (1) the test set, (2) subset of the test set with age below 65 years, (3) subset of the test set with age 65 and above, (4) female subset of the test set, and (5) male subset of the test set. To perform external validation, we applied the models to identical settings across the other available data sources not used for development of the model: LPD France, DA Germany, IPCI Netherlands, and SIDIAP CMBD-AH from Spain. Evaluation in the external data source was done on the full data set, since no train-test split was required.

Importantly, we recalibrated the models for each external database using mean calibration (or also referred to as calibration-in-the-large). Mean calibration adjusts the average predicted risk to the observed event rate by using a correction factor based upon the difference between those two. In addition the intercept is adjusted. As a result, if we were to underpredict the risk in the external population, the factor will increase the risk by some amount, and vice versa for overprediction [Van Calster B et al. BMC Med 2019].

These analyses were conducted and reported according to the Transparent Reporting of a multivariate prediction model for Individual Prediction or Diagnosis (TRIPOD) guidelines and adhered to the open science principles for publicly pre-specifying and tracking changes to study objects, protocol and code as described in the book of OHDSI.

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 ran cohort diagnostics package (associated analytic code https://github.com/oxford-pharmacoepi/CovCoagExposureDiagnostics and 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 step-by-step quality controls were followed for defining these cohorts as described in a previous progress report.

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, in their source coding systems, and in their linkage availability. Not all outcomes were available in all databases, given the context where some of the rarest events are treated and diagnosed. For example, mortality was not accurately recorded in data from DE and FR, precluding competing risk analyses in these countries.

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 has been written and will be submitted in the coming days, focussing on the more relevant aspects of our analyses.

13 Study Results

13.1 Baseline Characteristics

The number of subjects available in the primary study cohorts (diagnosed with COVID-19 or with a positive PCR test result) were as follows: 439,958 in CPRD UK, 45,370 in DA DE, 43,151 in IPCI NL, 103,086 in LPD FR, and 484,810 in SIDIAP CMBD-AH ES.

There was a predominance of female participants in all databases, and median (IQR) age ranged from 43 (25-58) in ES to 47 (30-60) in DE.

All contributing data sources provided data from March 2020, and run until January 2021 (UK), March 2021 (DE, NL), April 2021 (FR), and June 2021 (ES). They also provided a long period of observation before inclusion, with a median (IQR) ranging from 6.7 (3.7-12.2) years in DE to 14.8 (13.4-15.0) years in ES.

Detailed characteristics are reported in **Table 1**. Participants were generally healthy, with hypertension (13.4% in NL to 31.1% in DE), asthma (6.2% in ES to 12.4% in DE) and diabetes mellitus (6.8% in FR to 13.0 in DE%) being the 3 most prevalent comorbidities. Smoking was likely better recorded in primary care databases, with prevalence estimates of ever smokers of 10.5% (NL), 17.7% (ES) and 23.9% (UK) compared to only 1.5% (FR) and 3.3% (DE) in other data sources. NSAIDs were the most common medicines, present in 2.8% (UK) to 15.4% (ES) of the participating cohorts.

Table 1 - Baseline characteristics of primary study population (diagnosed with COVID-19 or with a positive PCR test)

Variable	CPRD AURUM UK	DA Germany	IPCI NL	LPD France	SIDIAP CMBDH-AH ES
N	439,958	45,370	43,151	103,086	484,810
Study period start	2020-03-01	2020-03-02	2020-03-01	2020-03-06	2020-03-01
Study period end	2021-01-30	2021-03-31	2021-03-03	2021-04-30	2021-06-22
Age – Median [IQR]	43 [31 to 56]	47 [30 to 60]	47 [29 to 59]	44 [30 to 57]	43 [25 to 58]
Age: Under 20	14,450 (3.3%)	4,761 (10.5%)	5,340 (12.4%)	11,192 (10.9%)	87,651 (18.1%)
Age: 20 to 29	86,236 (19.6%)	6,034 (13.3%)	5,957 (13.8%)	13,872 (13.5%)	60,362 (12.5%)
Age: 30 to 39	85,773 (19.5%)	6,810 (15.0%)	5,436 (12.6%)	16,930 (16.4%)	66,336 (13.7%)
Age: 40 to 49	82,801 (18.8%)	7,142 (15.7%)	7,178 (16.6%)	20,094 (19.5%)	86,275 (17.8%)
Age: 50 to 59	83,293 (18.9%)	8,906 (19.6%)	8,653 (20.1%)	19,850 (19.3%)	70,889 (14.6%)
Age: 60 to 69	44,540 (10.1%)	5,248 (11.6%)	5,329 (12.3%)	11,383 (11.0%)	43,962 (9.1%)
Age: 70 to 79	21,210 (4.8%)	2,674 (5.9%)	3,157 (7.3%)	6,374 (6.2%)	31,525 (6.5%)
Age: 80 or older	21,655 (4.9%)	3,795 (8.4%)	2,101 (4.9%)	3,391 (3.3%)	37,810 (7.8%)
Comorbidities					
Sex: Male	197,275 (44.8%)	20,641 (45.5%)	19,196 (44.5%)	44,999 (43.7%)	228,397 (47.1%)

Smoking	105,281 (23.9%)	1,509 (3.3%)	4,522 (10.5%)	1,496 (1.5%)	85,889 (17.7%)
Years of prior observation time Median [IQR]	12.0 [4.6 to 22.6]	6.7 [3.7 to 12.2]	6.8 [4.8 to 9.6]	7.2 [3.5 to 9.1]	14.8 [13.4 to 15.0]
Comorbidities (any time prior)					
Autoimmune disease	7,708 (1.8%)	3,022 (6.7%)	501 (1.2%)	1,301 (1.3%)	8,287 (1.7%)
Antiphospholipid syndrome	225 (0.1%)	<5	<5	<5	247 (0.1%)
Thrombophilia	520 (0.1%)	102 (0.2%)	<5	24 (0.0%)	555 (0.1%)
Asthma	50,796 (11.5%)	5,622 (12.4%)	3,130 (7.3%)	10,948 (10.6%)	30,072 (6.2%)
COPD	8,402 (1.9%)	4,162 (9.2%)	940 (2.2%)	1,727 (1.7%)	15,323 (3.2%)
Dementia	7,586 (1.7%)	1,885 (4.2%)	334 (0.8%)	275 (0.3%)	13,594 (2.8%)
Malignant neoplastic disease	18,426 (4.2%)	3,423 (7.5%)	2,910 (6.7%)	3,264 (3.2%)	31,634 (6.5%)
Diabetes mellitus	36,212 (8.2%)	5,910 (13.0%)	3,138 (7.3%)	7,049 (6.8%)	43,022 (8.9%)
Obesity	16,907 (3.8%)	6,234 (13.7%)	1,313 (3.0%)	1,240 (1.2%)	89,618 (18.5%)
Heart disease	30,931 (7.0%)	10,138 (22.3%)	3,728 (8.6%)	5,800 (5.6%)	59,105 (12.2%)
Hypertensive disorder	61,919 (14.1%)	14,115 (31.1%)	5,791 (13.4%)	16,810 (16.3%)	89,204 (18.4%)
Atrial fibrillation	9,088 (2.1%)	1,110 (2.4%)	1,032 (2.4%)	242 (0.2%)	15,959 (3.3%)
Renal impairment	20,118 (4.6%)	2,491 (5.5%)	1,224 (2.8%)	781 (0.8%)	28,337 (5.8%)
Medication use (prior 6 months)					
Non-steroidal anti-inflammatory drugs	12,243 (2.8%)	6,329 (13.9%)	3,629 (8.4%)	14,115 (13.7%)	74,681 (15.4%)
Cox2 inhibitors	308 (0.1%)	556 (1.2%)	251 (0.6%)	400 (0.4%)	2,637 (0.5%)
Systemic corticosteroids	6,456 (1.5%)	1,109 (2.4%)	1,735 (4.0%)	5,683 (5.5%)	12,996 (2.7%)
Antithrombotic and anticoagulant therapies	5,955 (1.4%)	2,347 (5.2%)	1,691 (3.9%)	4,748 (4.6%)	11,901 (2.5%)
Lipid modifying agents	10,901 (2.5%)	2,247 (5.0%)	1,775 (4.1%)	5,445 (5.3%)	9,668 (2.0%)
Antineoplastic and immunomodulating agents	3,101 (0.7%)	286 (0.6%)	2,049 (4.7%)	676 (0.7%)	4,832 (1.0%)
Hormonal contraceptives for systemic use	11,787 (2.7%)	463 (1.0%)	1,954 (4.5%)	3,285 (3.2%)	6,464 (1.3%)
Tamoxifen	106 (0.0%)	20 (0.0%)	36 (0.1%)	23 (0.0%)	148 (0.0%)
Sex hormones and modulators of the genital system	16,322 (3.7%)	583 (1.3%)	2,389 (5.5%)	4,236 (4.1%)	7,943 (1.6%)
Immunoglobulins	<5	<5	12 (0.0%)	8 (0.0%)	296 (0.1%)
Antithrombotics	10,133 (2.3%)	3,363 (7.4%)	3,159 (7.3%)	3,529 (3.4%)	32,682 (6.7%)
Corticosteroids	2,631 (0.6%)	514 (1.1%)	526 (1.2%)	463 (0.4%)	7,209 (1.5%)
Coxibs / Cox-2 inhibitors	216 (0.0%)	152 (0.3%)	142 (0.3%)	46 (0.0%)	1,296 (0.3%)
Immunoglobulins	<5	<5	<5	<5	40 (0.0%)
Lipid modifying agents	20,525 (4.7%)	3,198 (7.0%)	3,992 (9.3%)	4,159 (4.0%)	35,518 (7.3%)
Hormonal contraceptives	11,047 (2.5%)	344 (0.8%)	2,158 (5.0%)	1,108 (1.1%)	11,854 (2.4%)
Sex hormones modulators	71 (0.0%)	21 (0.0%)	25 (0.1%)	655 (0.6%)	61 (0.0%)

Baseline characteristics for people included in additional cohorts and in sensitivity analyses are available in an interactive web-app under the 'Patient profiles' tab. One interesting (although

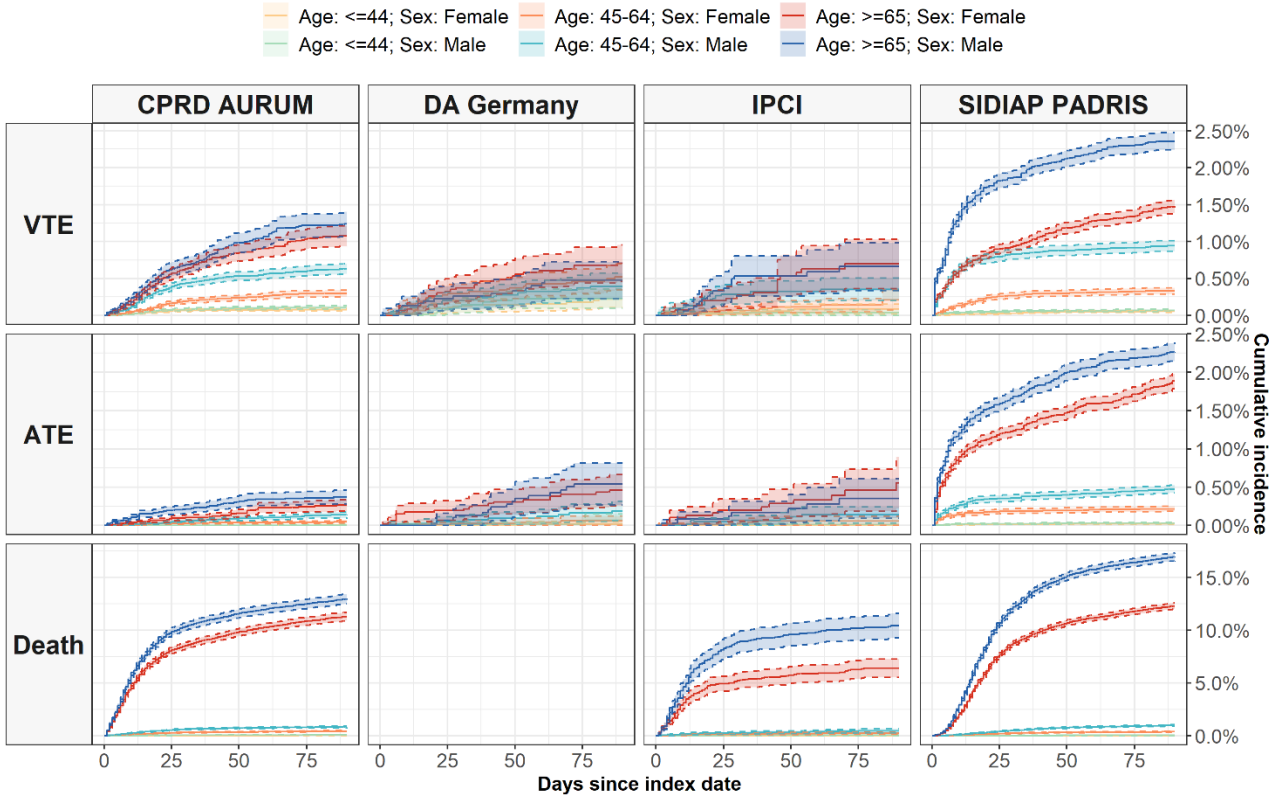
unsurprising) finding was the older age and male predominance of the hospitalized cohort in SIDIAP CMBD-AH, and the related high prevalence of comorbidities of interest in this population.

13.2 Cumulative incidence of VTE (Objective 4)

The overall cumulative incidence of DVT, PE and VTE in the primary cohorts for each database are detailed in **Table 2**. Sensitivity analyses for additional cohorts and without wash-out, and cumulative incidence of additional outcomes are reported in **Appendix 1**.

Age and sex-stratified cumulative incidence figures for VTE, ATE and death are depicted in **Figure 3**. All four outcomes were more common in men vs women, and in older age strata.

Figure 3 - Cumulative incidence of VTE, ATE, and all-cause mortality up to 90 days after a PCR+ or clinical diagnosis of COVID. CIF is reported for CPRD AURUM, IPCI, and SIDIAP CMBD-AH; KM is reported for DA Germany. LPD Fr was excluded from VTE analyses due to previously described under-reporting of these events



Overall 30-, 60-, and 90-day cumulative incidences of DVT ranged from 0.05%, 0.07%, 0.09% (UK) to 0.12%, 0.16%, 0.19% (ES). PE was more common, with equivalent figures ranging from 0.10%, 0.14%, 0.16% (NL) to 0.46%, 0.52%, 0.54% (UK). The cumulative incidence of VTE ranged from 0.13%, 0.20%, 0.24% (NL) to 0.54%, 0.62%, 0.66% (ES). As depicted in Figure 3, the cumulative incidence of VTE was higher in men vs women and increased with older age.

Table 2 - Cumulative incidence of DVT, PE, and VTE

Database	Outcome	Time from index (days)	Number at risk at end of interval	Events during interval	Cumulative incidence (95%CI)	Estimator
CPRD AURUM	DVT	30	330,815	194	0.05% (0.05% to 0.04%)	CIF
CPRD AURUM	DVT	60	179,167	263	0.07% (0.08% to 0.07%)	CIF
CPRD AURUM	DVT	90	119,798	297	0.09% (0.11% to 0.08%)	CIF
DA Germany	DVT	30	30,313	17	0.05% (0.03% to 0.07%)	KM
DA Germany	DVT	60	26,446	29	0.09% (0.06% to 0.13%)	KM
DA Germany	DVT	90	22,508	38	0.13% (0.09% to 0.17%)	KM
IPCI	DVT	30	31,335	19	0.05% (0.08% to 0.03%)	CIF
IPCI	DVT	60	22,090	28	0.08% (0.12% to 0.05%)	CIF
IPCI	DVT	90	11,347	30	0.09% (0.13% to 0.06%)	CIF
SIDIAP CMBD-AH	DVT	30	509,562	706	0.12% (0.13% to 0.11%)	CIF
SIDIAP CMBD-AH	DVT	60	472,498	937	0.16% (0.17% to 0.15%)	CIF
SIDIAP CMBD-AH	DVT	90	431,770	1,059	0.19% (0.20% to 0.18%)	CIF
CPRD AURUM	PE	30	330,084	817	0.19% (0.21% to 0.18%)	CIF
CPRD AURUM	PE	60	178,712	949	0.24% (0.26% to 0.23%)	CIF
CPRD AURUM	PE	90	119,479	998	0.27% (0.29% to 0.25%)	CIF
DA Germany	PE	30	30,231	60	0.18% (0.13% to 0.22%)	KM
DA Germany	PE	60	26,364	85	0.27% (0.21% to 0.32%)	KM
DA Germany	PE	90	22,437	93	0.30% (0.24% to 0.36%)	KM
IPCI	PE	30	31,369	38	0.10% (0.14% to 0.07%)	CIF
IPCI	PE	60	22,110	48	0.14% (0.18% to 0.10%)	CIF
IPCI	PE	90	11,358	52	0.16% (0.21% to 0.12%)	CIF
SIDIAP CMBD-AH	PE	30	507,897	2,647	0.46% (0.48% to 0.44%)	CIF
SIDIAP CMBD-AH	PE	60	471,434	2,978	0.52% (0.54% to 0.50%)	CIF
SIDIAP CMBD-AH	PE	90	430,907	3,099	0.54% (0.56% to 0.52%)	CIF
CPRD AURUM	VTE	30	329,421	969	0.23% (0.24% to 0.21%)	CIF
CPRD AURUM	VTE	60	178,204	1,165	0.31% (0.32% to 0.29%)	CIF
CPRD AURUM	VTE	90	119,079	1,236	0.35% (0.37% to 0.33%)	CIF
DA Germany	VTE	30	30,109	74	0.22% (0.17% to 0.27%)	KM
DA Germany	VTE	60	26,240	109	0.34% (0.28% to 0.41%)	KM

DA Germany	VTE	90	22,316	125	0.41% (0.34% to 0.48%)	KM
IPCI	VTE	30	31,258	52	0.13% (0.18% to 0.10%)	CIF
IPCI	VTE	60	22,025	71	0.20% (0.26% to 0.16%)	CIF
IPCI	VTE	90	11,308	77	0.24% (0.30% to 0.19%)	CIF
SIDIAP CMBD-AH	VTE	30	506,470	3,117	0.54% (0.56% to 0.52%)	CIF
SIDIAP CMBD-AH	VTE	60	470,262	3,585	0.62% (0.64% to 0.60%)	CIF
SIDIAP CMBD-AH	VTE	90	429,815	3,795	0.66% (0.68% to 0.64%)	CIF

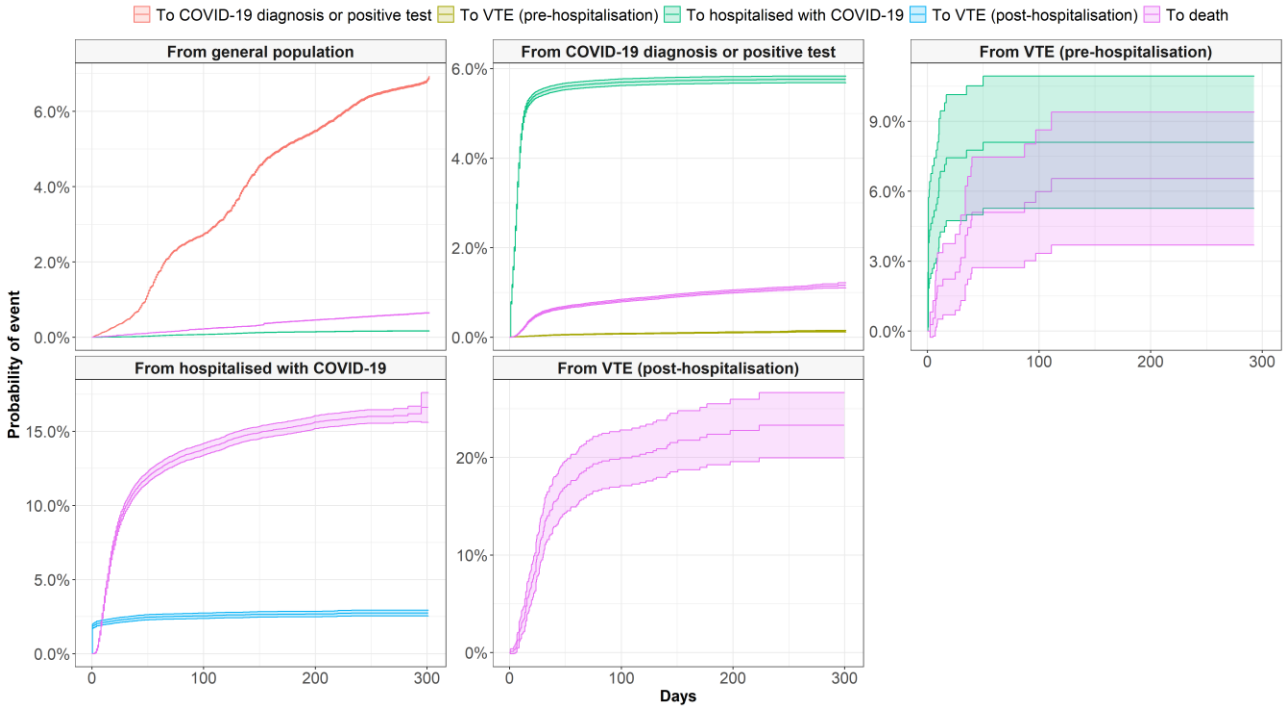
13.3 The association between VTE and COVID-19 worsening (Objective 5)

A population of 5,489,533 people from the general population entered the multistate model as of 1st September 2020. Of these, 376,349 entered the COVID-19 diagnosis or test positive state. Among these people, 370 with VTE prior to a COVID-19 hospitalisation were identified. Of the 30,219 persons hospitalised with COVID-19, 783 had a VTE identified on the day of admission or subsequently. Baseline characteristics for participants at each state are reported in **Appendix 2**.

The cumulative incidence for each of the transitions in the model is shown below in **Figure 4**. Over the study period, the cumulative incidence of COVID-19 diagnosis or positive test was 6.9%, with a further 0.2% for the direct transition from general population to hospitalised with COVID-19. The 90-day cumulative incidence of VTE prior to hospiatlisation among those diagnosed with COVID-19 or with a positive test was 0.07%, while the cumulative incidence of VTE following hospitalisation was 2.5%.

After adjustment for age and sex, VTE prior to COVID-19 hospitalisation was associated with a HR of 1.36 (95% CI: 0.95 to 1.96) and 4.42 (3.07 to 6.36) for death without hospitalisation. VTE on or after date of hospital admission was associated with a HR of 1.63 (1.39 to 1.90) for death.

Figure 4 - Cumulative incidence of transitions in VTE multistate model



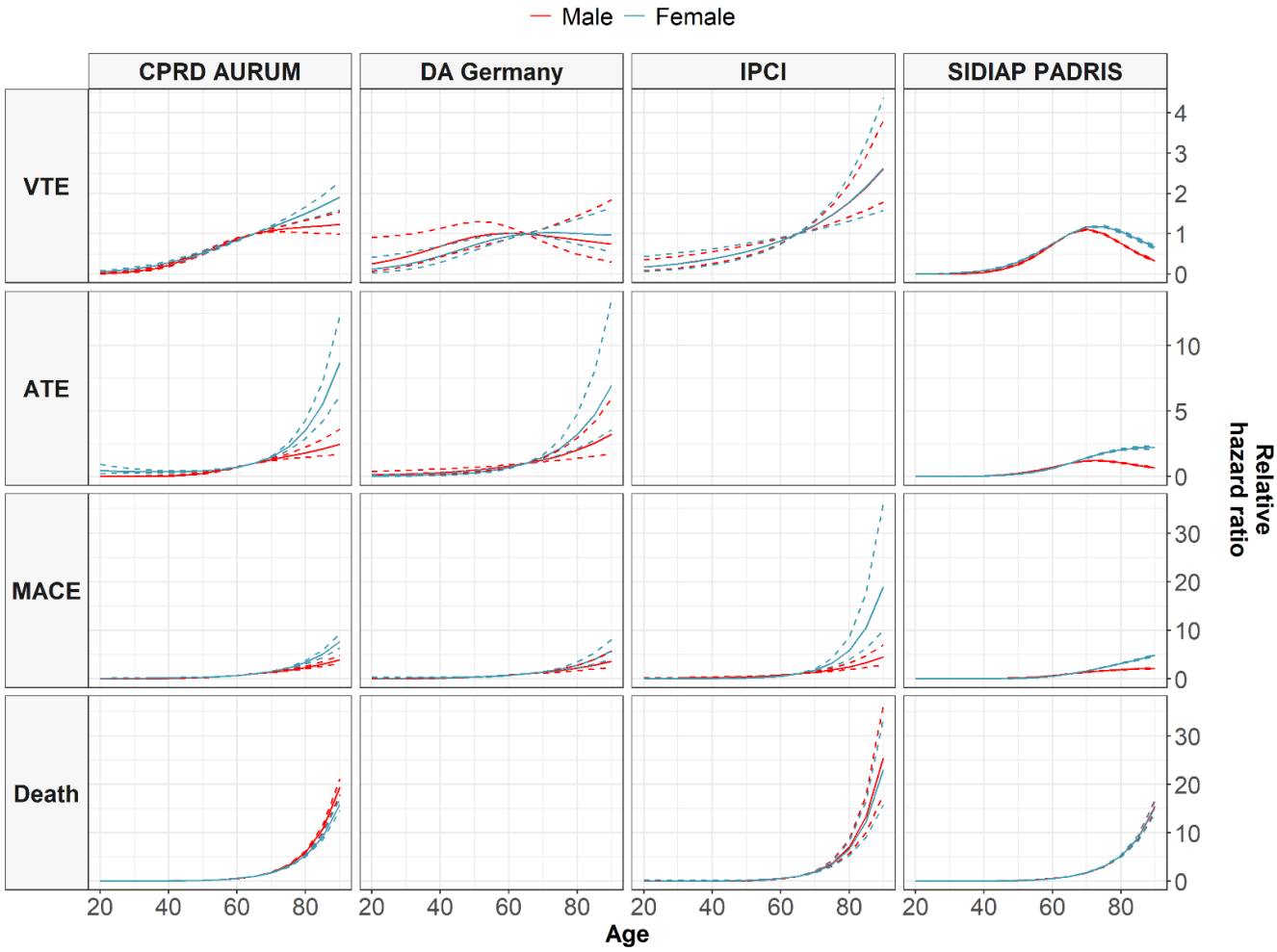
13.4 Pre-specified risk factors for VTE (Objective 6)

Survival models for the association between sex and VTE in COVID-19 patients showed a higher risk of VTE in men compared to women, with age-adjusted HR ranging from 0.79 (0.55 to 1.14) in DA DE to 1.44 (1.41 to 1.48) in SIDIAP CMBD-AH ES. DVT had a less consistent association with sex, with a lower risk in men vs women in ES (age-adjusted HR 0.90 (0.85 to 0.94)) and no association with sex in UK data (age-adjusted HR 1.18 (0.96 to 1.45)). PE was consistently more common in men than women, with age-adjusted HRs 1.21 (1.08 to 1.35) in the UK, 2.14 (1.34 to 3.42) in the NL, and 1.54 (1.49 to 1.59) in ES. Less clear signals were observed in DE and FR, with age-adjusted HR 0.74 (0.48 to 1.13) and 1.09 (0.75 to 1.59) respectively. Sensitivity analyses including additional cohorts, and the result of the association between sex and additional outcomes including MACE and all-cause mortality are reported in full in **Appendix 3** and in the accompanying web app in the 'Hazard Ratios for sex' tab.

Sex-stratified HR (95%CI) for age in association with all the study outcomes (including the additional MACE and death) are depicted in **Figure 5**. As observed, the association between age and VTE appeared to take an inverse U-shape in most databases (UK, DE, and ES): risk increased with age from 40 to 70, to then decline with older ages above 70 years old.

Full results for the association between age and all study outcomes, including additional cohorts and sensitivity analyses, are reported in **Appendix 4**.

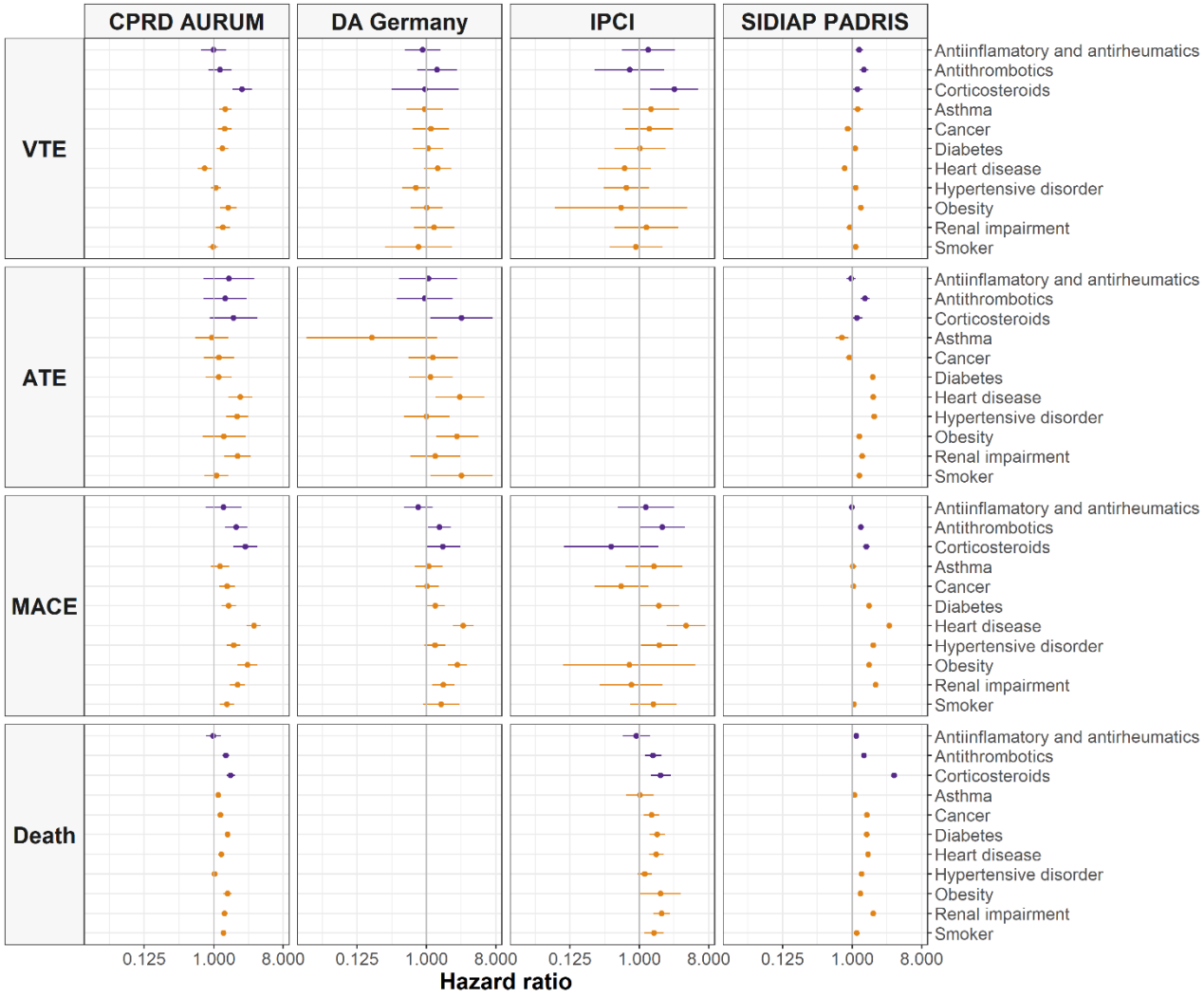
Figure 5 - Hazard ratios (95%CI) for the association between age and VTE, ATE, MACE and Death. NOTE: too few ATE events were seen in IPCI, precluding modelling. Death was incomplete in DA Germany.



Age and sex-adjusted HR (95%CI) for all other pre-specified risk factors are depicted in **Figure 6**. Full results including sensitivity analyses, additional cohorts and outcomes, and both cause-specific as well as non-informative Cox regression models are reported in **Appendix 5**.

The comorbidities associated with DVT included asthma (adjusted HR 1.71 (1.32-2.22) in UK data), cancer (adjusted HR 1.89 (1.46-2.45) in UK data), hypertension (adjusted HR 1.34 (1.08-1.67) in the UK, 1.48 (1.40-1.57) in ES), renal impairment (adjusted HR 1.62 (1.24-2.11) in the UK), and obesity (adjusted HR 2.40 (1.73-3.34) in UK, HR 1.28 (1.21-1.35) in ES). Smoking appeared also associated with an increased risk of DVT (adjusted HR 1.44 (1.37 to 1.53) in ES). As for medicines, the use of antithrombotics (adjusted HR 2.30 (1.61-3.27) in the UK, HR 1.31 (1.21-1.42) in ES) and systemic corticosteroids (adjusted HR 1.53 (0.98-2.41) in the UK, 1.10 (1.01-1.20) in ES) were associated with an excess risk of DVT in some databases.

Figure 6 - Adjusted HR for the association between comorbidity, medicines use and the risk of VTE, ATE, MACE, Death



As for PE, asthma appeared associated with an increased risk in UK data (adjusted HR 1.42 (1.22-1.65)), diabetes mellitus in the UK (adjusted HR 1.19 (1.03-1.37)) and NL (adjusted HR 2.53 (1.51-4.23), malignancy in the UK (adjusted HR 1.42 (1.21-1.67)), obesity in UK (adjusted HR 1.70 (1.39-2.08)), and renal impairment in UK (adjusted HR 1.38 (1.18-1.63)) and NL (adjusted HR 4.01 (2.23-7.22)). Medicines associated with an increased risk of PE included systemic glucocorticoids (adjusted HR 2.99 (2.45-3.64) in UK, 2.79 (1.53-5.07) in NL) and NSAIDs albeit only in NL data (adjusted HR 1.99 (1.09-3.61)).

Finally, VTE appeared associated with the following comorbidities: asthma (adjusted HR 1.50 (1.31-1.71) in UK), diabetes mellitus (adjusted HR 1.21 (1.06-1.38) in UK), malignancy (adjusted HR 1.60 (1.39-1.84) in the UK, and 1.61 (0.95-2.75) in the NL), obesity (adjusted HR 1.89 (1.59-2.25) in UK), and renal impairment (adjusted HR 1.40 (1.21-1.61) in the UK, 2.43 (1.29-4.59) in the NL). Previous use of antithrombotics (adjusted HR 1.40 (1.12-1.76) in the UK), NSAIDs

(adjusted HR 2.06 (1.23-3.45) in the NL), and systemic glucocorticoids (adjusted HR 2.73 (2.27-3.28) in the UK, 2.96 (1.73-5.08) in the NL) appeared associated with an increased risk of VTE.

13.5 Algorithms for the prediction of VTE in COVID-19 patients (Objective 7)

Models were developed for VTE and its subcomponents (DVT and PE) separately. Baseline (age-sex only) and parsimonious (age, sex and pre-specified risk factors) achieved similar performance for all VTE and ATE outcomes in internal validation, as demonstrated in **Figure 7**.

Figure 7 - Internal validation performance in CPRD AURUM: AUROC discrimination (top) and Eavg calibration performance (bottom). AUROC is reported with 95% Confidence Intervals.

As shown, baseline models for VTE had good discrimination (AUC 0.80), and the addition of pre-specified risk factors did not improve this despite increasing complexity. Discrimination was similar in men (AUC 0.77) vs women (AUC 0.82) and in COVID-19 patients aged <65 (AUC 0.80), but declined quite significantly in older ages >=65 (AUC 0.52). Discrimination was better for PE (AUC 0.77 in baseline, 0.78 in parsimonious models) than for DVT (AUC 0.67 and 0.66). Calibration was good as demonstrated by low Eavg values in **Figure 7**, and good correlation between expected and observed risk in most age-sex strata (**Figure 8**).

The full parsimonious model coefficients are reported in **Table 3** for further implementation and to maximize the use and reproducibility of this work.

Figure 8 - Age and sex-specific observed (red dash) vs expected (blue solid) risk of PE, DVT, VTE

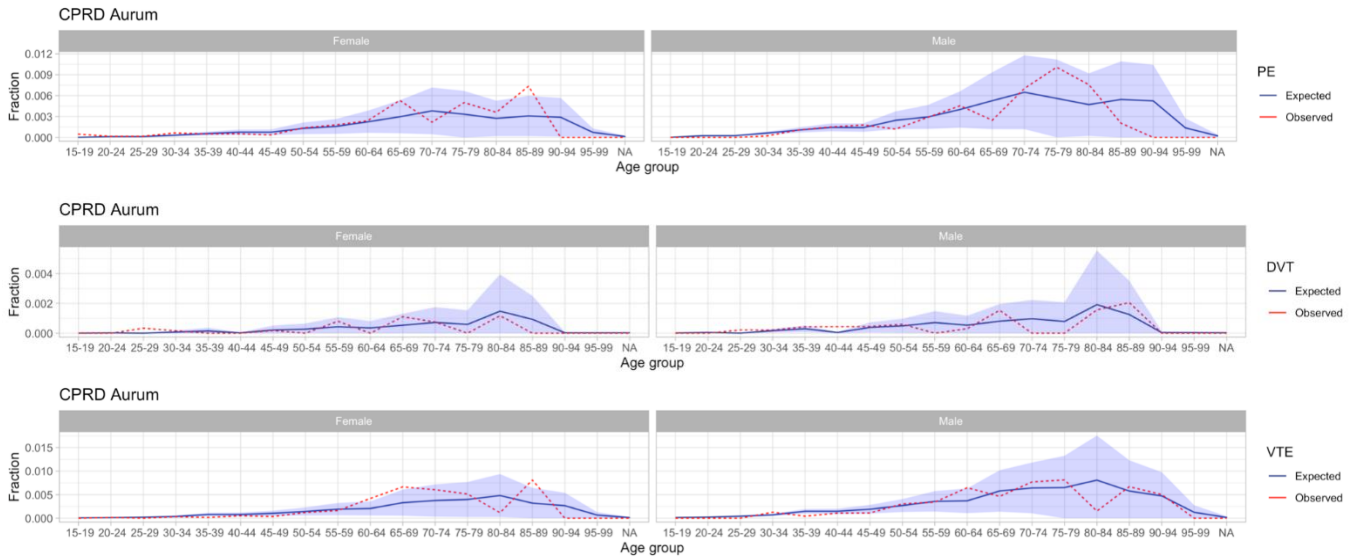


Table 3 - Parameters in the parsimonious LASSO models for 30-day PE, DVT, and VTE

	PE at 30 days	**DVT at 30 days**	**VTE at 30 days**
Intercept	-8.965058618	-10.33370024	-8.98444339
Sex, male	0.607009816	0.38782297	0.59974911
Autoimmune disease	0.374821645	-0.04848423	0.38242008
Atrial fibrillation	-1.017169177	-3.48202393	-1.43145865
Asthma	0.245977181	0.48931976	0.21448469
Thrombophilia	-0.893965181	0	-0.87944210
Antiphospholipid syndrome	0	0	0
Malignant neoplastic disease	0.208920469	-0.02111686	0.20675827
Diabetes mellitus	-0.110404538	0.23588421	0.02852389
Obesity	0.481353761	0.40999818	0.38067695
Heart disease	-0.121245055	-0.30215219	-0.34315253
Hypertensive disorder	0.058797845	-0.31644568	0.12609758
Renal impairment	0.216441503	0.12788905	0.27794847
Chronic obstructive lung disease	0.580501283	-1.63435595	0.30090704

Dementia	0.095582574	-0.35118043	0.10358257
Antiinflammatory	0.242091467	0.19846435	0.20980817
Coxib	-0.654695322	-1.80801286	-0.79944730
Corticosteroids	-0.040170883	0.62478390	0.10465208
Antithrombotic agent	-0.000435282	0.37029871	-0.08458989
Lipid modifying agent	0.197905448	0.22573582	0.22720536
Antineoplastic immunomodulating	0	-0.10976097	0.10268507
Hormonal contraceptives	0.058134589	0	0.20840583
Tamoxifen	-1.349590423	0	-1.56735980
Sex hormones modulators	0.311349203	-1.53647223	-0.36876343
Age 15-19	-2.242227161	-1.07747299	-0.48640652
Age 20-24	0	0	0
Age 25-29	0.005851790	-2.13777315	0.54367601
Age 30-34	0.908224777	1.12544758	1.05903629
Age 35-39	1.425317213	1.67903250	1.78076669
Age 40-44	1.715427384	-0.13089396	1.76171740
Age 45-49	1.672746453	1.90127506	1.99300634
Age 50-54	2.206268155	2.12307407	2.30631484
Age 55-59	2.358760108	2.49598708	2.56869361
Age 60-64	2.642220392	2.20005357	2.56911834
Age 65-69	2.882434988	2.58996535	2.98020591
Age 70-74	3.074052792	2.87688970	3.08959731
Age 75-79	2.910722768	2.73407206	3.13038577
Age 80-84	2.766674981	3.73278311	3.40153928
Age 85-89	2.967961077	3.40744124	3.05357582
Age 90-94	2.957189681	0	2.93820878
Age 95-99	1.598737161	0	1.62547340
Age 100-104	0	0	0
Age 105-109	0	0	0

External validation of both baseline and parsimonious models are reported in **Figure 9a** (IPCI, NL), **9b** (SIDIAP, ES), **9c** (LPD FR) and **9d** (DA Germany). Results were similar to those seen in the internal validation except for DA Germany, were discrimination and calibration were worse than in the UK data. Low power due to previously reported issues with the coding of VTE and smaller cohort size in LPD FR and DA DE limited our ability to validate in different age strata. DVT models could not be validated in LPD FR due to similar reasons.

Figure 9 - External validation performance: AUROC discrimination and Eavg calibration

Figure 9a. External validation results from IPCI (NL)

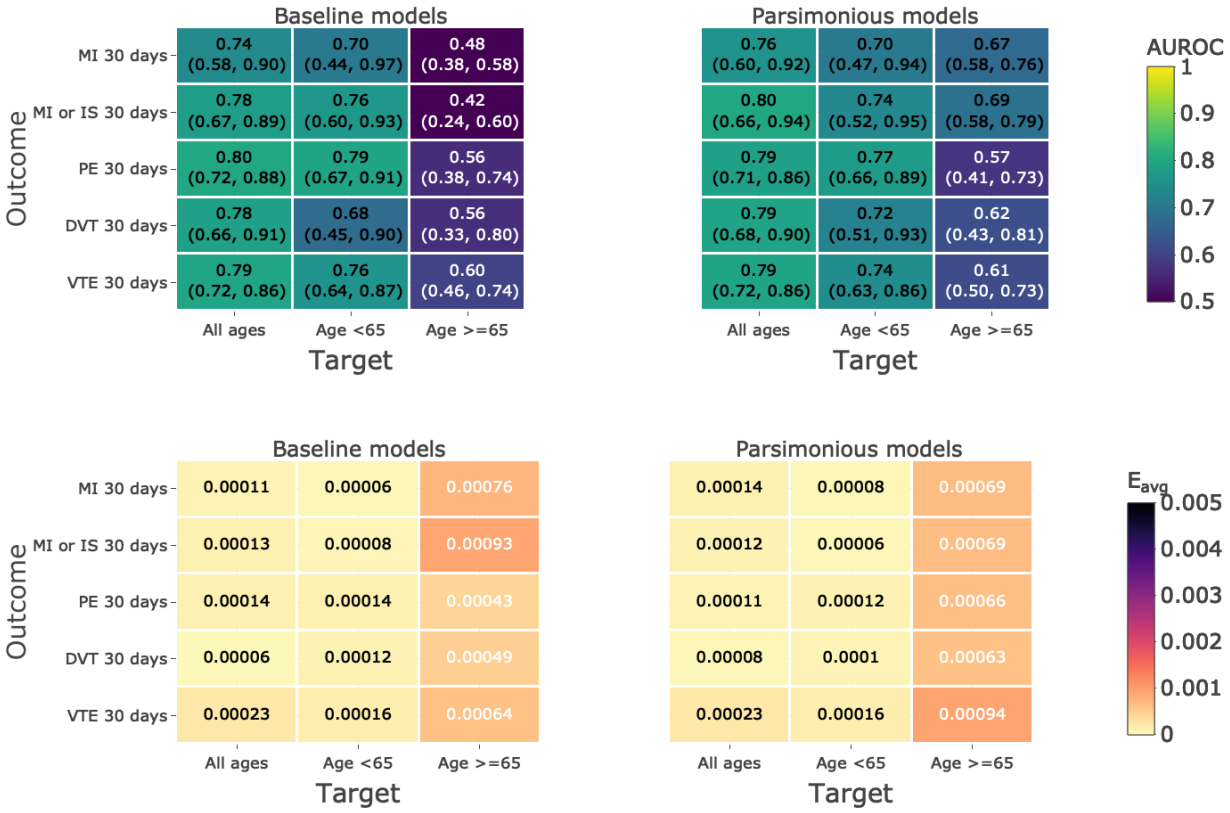


Figure 9b. External validation results from SIDIAP CMBD-AH (ES)

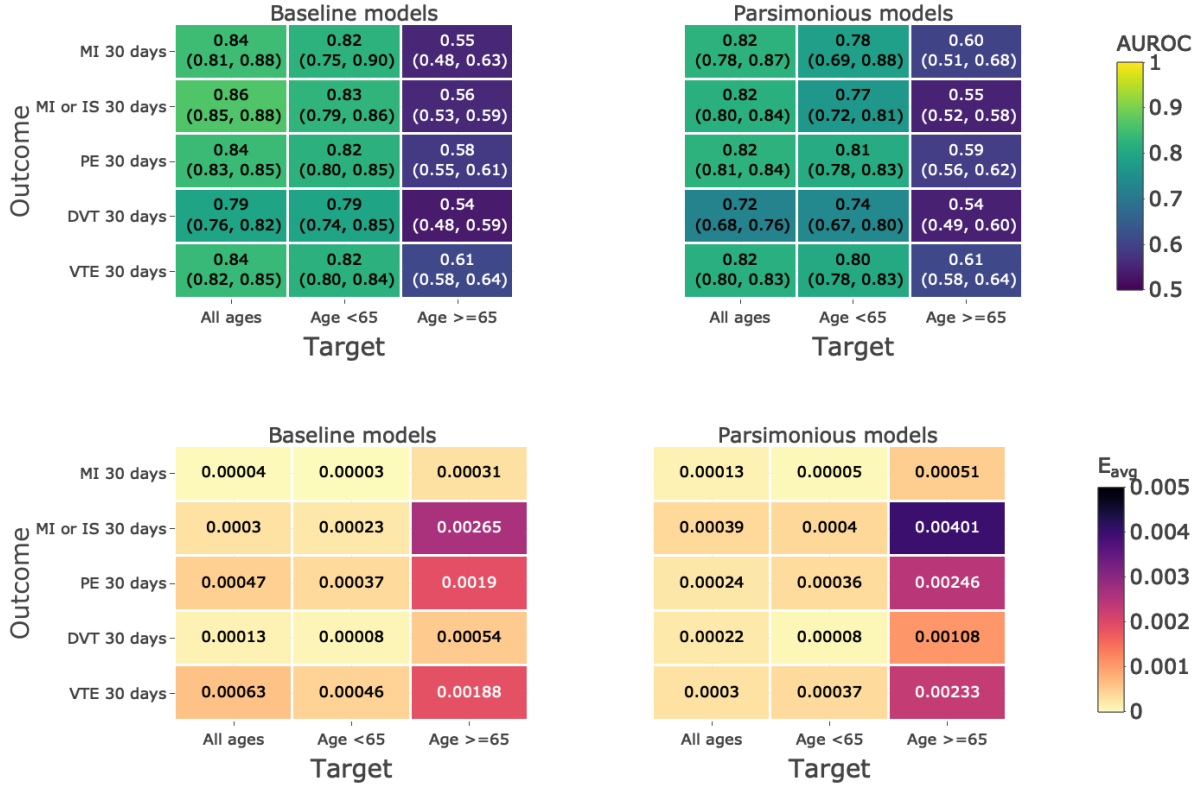
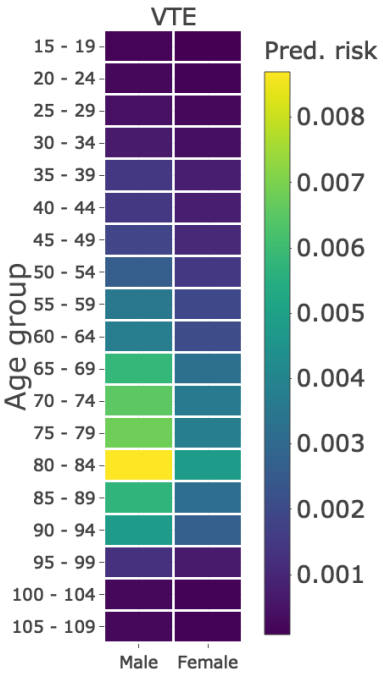


Figure 9c. External validation results from LPD (France)

Figure 9d. External validation results from DA (Germany)

The resulting simpler baseline (age-sex) models are shown in **Figure 10** in the form of a heatmap depicting predicted VTE risk for different age-sex strata. Caution is needed for the interpretation of risk predictions in COVID-19 patients aged >=65 due to the above reported decline in discrimination in older ages.

Figure 10 - Heatmap of predicted risk of VTE amongst COVID-19 patients



13.6 Cumulative incidence of ATE (Objective 8)

The overall cumulative incidence of ATE and its subcomponents Ischemic Stroke (IS) and Myocardial Infarction (MI) are reported in **Table 4**. Additional sensitivity analyses of additional outcomes, age-sex stratification, and additional cohorts are reported in **Appendix 1**.

IS was observed with great heterogeneity in COVID-19 patients, with 30-day cumulative incidence ranging from 0.01% (UK) to 0.49% (ES). MI risks also varied across databases, from a lowest 30-day cumulative incidence of 0.02% (FR) to 0.10% (ES). The 30-day cumulative incidence of ATE were (in ascending order): 0.03% (FR), 0.04% (UK), 0.05% (NL), 0.06% (DE), and much higher at 0.48% in ES. It is likely that such variability is at least partially attributable to previously described database heterogeneity related to coding practice, coverage, and linkage availability (with the data from ES having hospital linkage).

Table 4 - Cumulative incidence of IS, MI, and ATE

Database	Outcome	Time from index (days)	Number at risk at end of interval	Events during interval	Cumulative incidence (95%CI)	Estimator
CPRD AURUM	IS	30	331,358	36	0.01% (0.01% to 0.01%)	CIF
CPRD AURUM	IS	60	179,596	49	0.01% (0.02% to 0.01%)	CIF
CPRD AURUM	IS	90	120,151	53	0.02% (0.02% to 0.01%)	CIF
DA Germany	IS	30	30,283	15	0.04% (0.02% to 0.07%)	KM
DA Germany	IS	60	26,417	28	0.09% (0.06% to 0.12%)	KM
DA Germany	IS	90	22,482	36	0.12% (0.08% to 0.16%)	KM
IPCI	IS	30	31,393	<5	0.01% (0.02% to 0.00%)	CIF
IPCI	IS	60	22,139	7	0.02% (0.05% to 0.01%)	CIF
IPCI	IS	90	11,378	7	0.02% (0.05% to 0.01%)	CIF
SIDIAP CMBD-AH	IS	30	506,270	2,270	0.39% (0.41% to 0.38%)	CIF
SIDIAP CMBD-AH	IS	60	469,667	2,575	0.45% (0.47% to 0.43%)	CIF
SIDIAP CMBD-AH	IS	90	429,152	2,767	0.49% (0.50% to 0.47%)	CIF

CPRD AURUM	MI	30	330,689	126	0.03% (0.04% to 0.02%)	CIF
CPRD AURUM	MI	60	179,126	177	0.05% (0.06% to 0.04%)	CIF
CPRD AURUM	MI	90	119,823	195	0.06% (0.07% to 0.05%)	CIF
DA Germany	MI	30	30,300	7	0.02% (0.01% to 0.04%)	KM
DA Germany	MI	60	26,434	15	0.05% (0.02% to 0.07%)	KM
DA Germany	MI	90	22,500	22	0.08% (0.04% to 0.11%)	KM
IPCI	MI	30	31,245	21	0.05% (0.08% to 0.04%)	CIF
IPCI	MI	60	22,028	33	0.10% (0.14% to 0.07%)	CIF
IPCI	MI	90	11,317	38	0.13% (0.18% to 0.09%)	CIF
LPD France	MI	30	67,829	14	0.02% (0.01% to 0.03%)	KM
LPD France	MI	60	61,782	25	0.04% (0.02% to 0.05%)	KM
LPD France	MI	90	53,698	33	0.05% (0.03% to 0.07%)	KM
SIDIAP CMBD-AH	MI	30	509,804	552	0.10% (0.10% to 0.09%)	CIF
SIDIAP CMBD-AH	MI	60	472,784	687	0.12% (0.13% to 0.11%)	CIF
SIDIAP CMBD-AH	MI	90	432,048	767	0.13% (0.14% to 0.13%)	CIF
CPRD AURUM	ATE	30	330,398	160	0.04% (0.04% to 0.03%)	CIF
CPRD AURUM	ATE	60	178,942	219	0.06% (0.07% to 0.05%)	CIF
CPRD AURUM	ATE	90	119,691	241	0.08% (0.09% to 0.07%)	CIF
DA Germany	ATE	30	30,058	19	0.06% (0.03% to 0.08%)	KM
DA Germany	ATE	60	26,192	37	0.12% (0.08% to 0.16%)	KM
DA Germany	ATE	90	22,277	52	0.18% (0.13% to 0.23%)	KM
IPCI	ATE	30	31,121	19	0.05% (0.08% to 0.03%)	CIF
IPCI	ATE	60	21,949	29	0.09% (0.12% to 0.06%)	CIF
IPCI	ATE	90	11,275	33	0.11% (0.16% to 0.08%)	CIF
LPD France	ATE	30	67,460	13	0.02% (0.01% to 0.03%)	KM
LPD France	ATE	60	61,435	21	0.03% (0.02% to 0.04%)	KM

LPD France	ATE	90	53,386	29	0.04% (0.03% to 0.06%)	KM
SIDIAP CMBD-AH	ATE	30	504,636	2,738	0.48% (0.50% to 0.46%)	CIF
SIDIAP CMBD-AH	ATE	60	468,363	3,175	0.56% (0.58% to 0.54%)	CIF
SIDIAP CMBD-AH	ATE	90	427,980	3,445	0.61% (0.63% to 0.59%)	CIF

As observed in **Figure 3**, ATE was more common in men vs women, and in older vs younger ages.

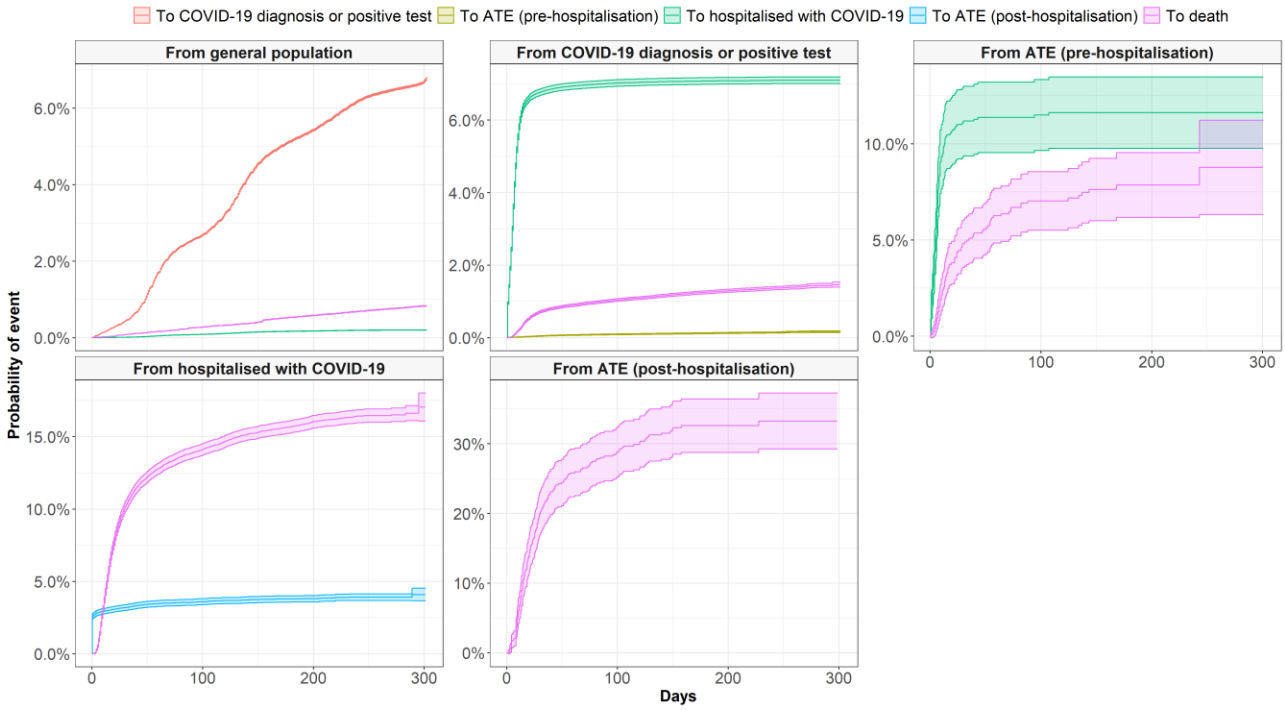
13.7 The association between ATE and COVID-19 worsening (Objective 9)

As with the models for VTE, a population of 5,489,533 people from the general population entered the multistate model for ATE as of 1st September 2020. Of these, 376,349 went on to enter the COVID-19 diagnosis or test positive state. Among these people, 1,181 with ATE prior to a COVID-19 hospitalisation were identified. Of the 30,219 persons hospitalised with COVID-19, 323 had an ATE identified on the day of admission or subsequently. Details on baseline characteristics at the time of entering each state are reported in **Appendix 2**.

The cumulative incidence for each of the transitions in the model is shown below in **Figure 11**. Over the study period, the cumulative incidence of COVID-19 diagnosis or positive test was 6.9%, with a further 0.2% for the direct transition from general population to hospitalised with COVID-19. The 90-day cumulative incidence of ATE prior to hospitalisation among those diagnosed with COVID-19 or with a positive test was 0.26%, while the cumulative incidence of ATE following hospitalisation was 0.9%.

After adjustment for age and sex, ATE prior to COVID-19 hospitalisation was associated with a HR of 1.05 (95% CI: 0.89 to 1.25) and 3.16 (2.65 to 3.75) for death without hospitalisation. ATE on or after date of hospital admission was associated with a HR of 1.93 (1.57 to 2.37) for death.

Figure 11 - Cumulative incidence of transitions in ATE multistate model



13.8 Pre-specified risk factors for ATE (Objective 10)

Cox models for the association between sex and ATE among COVID-19 patients showed a higher risk of ATE in men compared to women. Age-adjusted HR for the association between male sex and IS ranged from 0.69 (0.45 to 1.06) in the UK to 1.22 (1.18 to 1.27) in ES. Similar HR for the association between male sex and MI went from 1.23 (1.17 to 1.30) in ES to 2.09 (1.68 to 2.62) amongst UK patients. As for composite ATE, the HR for male sex went from 1.19 (1.15 to 1.23) in ES to 1.73 (1.42 to 2.10) amongst UK patients. Sensitivity analyses are reported in full in **Appendix 3** and in the accompanying web app in the 'Hazard Ratios for sex' tab.

Sex-stratified HR (95%CI) for age in association with ATE are depicted in **Figure 5**. As observed, the association between age and ATE takes an exponential shape, with weak increases in risk in ages 40-70 but rapid and higher effect sizes for the association between age and ATE risk in COVID-19 patients aged 70 or older. All results for the association between age and ATE, including sensitivity analyses, are reported in **Appendix 4** and in the accompanying app under the 'Relative hazard ratios for age' tab.

Age and sex-adjusted HRs (95%CI) for the pre-specified list of comorbidities and medicines use are depicted in **Figure 6**. All results including additional cohorts and sensitivity analyses are reported in **Appendix 5**. Comorbidities associated with IS include hypertension (adjusted HR 2.28 (1.44-3.59)

in the UK, 1.46 (1.40-1.52) in ES), diabetes mellitus (adjusted HR 1.34 (1.30-1.39) in ES), and heart disease history (adjusted HR 1.62 (1.46-1.69)). The only medicines use associated with IS were antithrombotics (adjusted HR 3.73 (1.98-7.01) in UK, 1.14 (1.08-1.20) in ES).

As for MI, the key comorbidities associated with an increased risk were diabetes mellitus (adjusted HR 1.80 (1.42-2.27) in UK, 1.50 (1.42-1.58) in ES), malignancy (adjusted HR 1.62 (1.25-2.11) in UK), heart disease (adjusted HR 2.35 (1.85-2.98) in the UK), hypertension (adjusted HR 4.14 (3.25-5.28) in the UK, 1.24 (1.17-1.31) in ES), obesity (adjusted HR 1.22 (1.15-1.29) in ES), and renal impairment (adjusted HR 4.54 (3.56-5.79) in UK data, 1.07 (1.01-1.13) in ES). Smoking appeared associated with an increased risk of MI, with an adjusted HR of 1.58 (1.25-2.00) in the UK data.

Finally, comorbidities associated with ATE amongst COVID-19 patients included diabetes mellitus (adjusted HR 1.65 (1.33-2.06) in the UK, 1.41 (1.37-1.45) in ES), history of heart disease (adjusted HR 2.01 (1.62-2.51) in UK, 2.74 (1.46-5.13) in DE, and 1.40 (1.35-1.44) in ES), hypertension (adjusted HR 4.13 (3.31-5.14) in UK, 1.39 (1.34-1.43) in ES), obesity (adjusted HR 3.09 (1.82-5.24) in DE, 1.05 (1.01-1.08) in ES), and renal impairment (adjusted HR 3.62 (2.89-4.53) in UK data, 1.87 (1.03-3.40) in DE, and 1.04 (1.01-1.07) in ES). Cancer appeared associated with ATE in UK data, with an adjusted HR of 1.46 (1.14-1.88). Smoking was consistently associated with an increased risk of ATE, with adjusted HR 1.41 (1.14-1.75) in the UK, and 2.53 (1.09-5.88) in DE. Previous antithrombotics use had an adjusted HR 1.80 (1.28-2.53) in the UK and 1.08 (1.04-1.14) in ES, and use of systemic glucocorticoids an adjusted HR of 5.28 (2.79-10.01) in DE.

13.9 Algorithms for the prediction of ATE in COVID-19 patients (Objective 11)

Models were generated for the prediction of ATE at 30-day periods following index date, and for MI separately. Models for IS did not converge, probably due to limited statistical power. Similar to the findings for VTE prediction, baseline (age-sex only) and parsimonious (age, sex and pre-specified risk factors in the previous section) models performance similarly for the prediction of ATE, MI and IS, as demonstrated in **Figure 7**. Discrimination was excellent for the prediction of ATE in people <65 in internal validation data from the UK (AUC 0.85 for MI, 0.84 for ATE), but - similar to VTE results- it declined significantly for patients aged >=65 (AUC 0.55 for MI, 0.58 for ATE). This latter model (for the prediction of ATE in ages>=65) was the only where discrimination increased to acceptable levels with the addition of risk factors in the parsimonious model, with AUC increasing from 0.58 in the baseline (age-sex only) algorithm to 0.73 in the more complex one. The coefficients in the resulting parsimonious models are reported in **Table 5**.

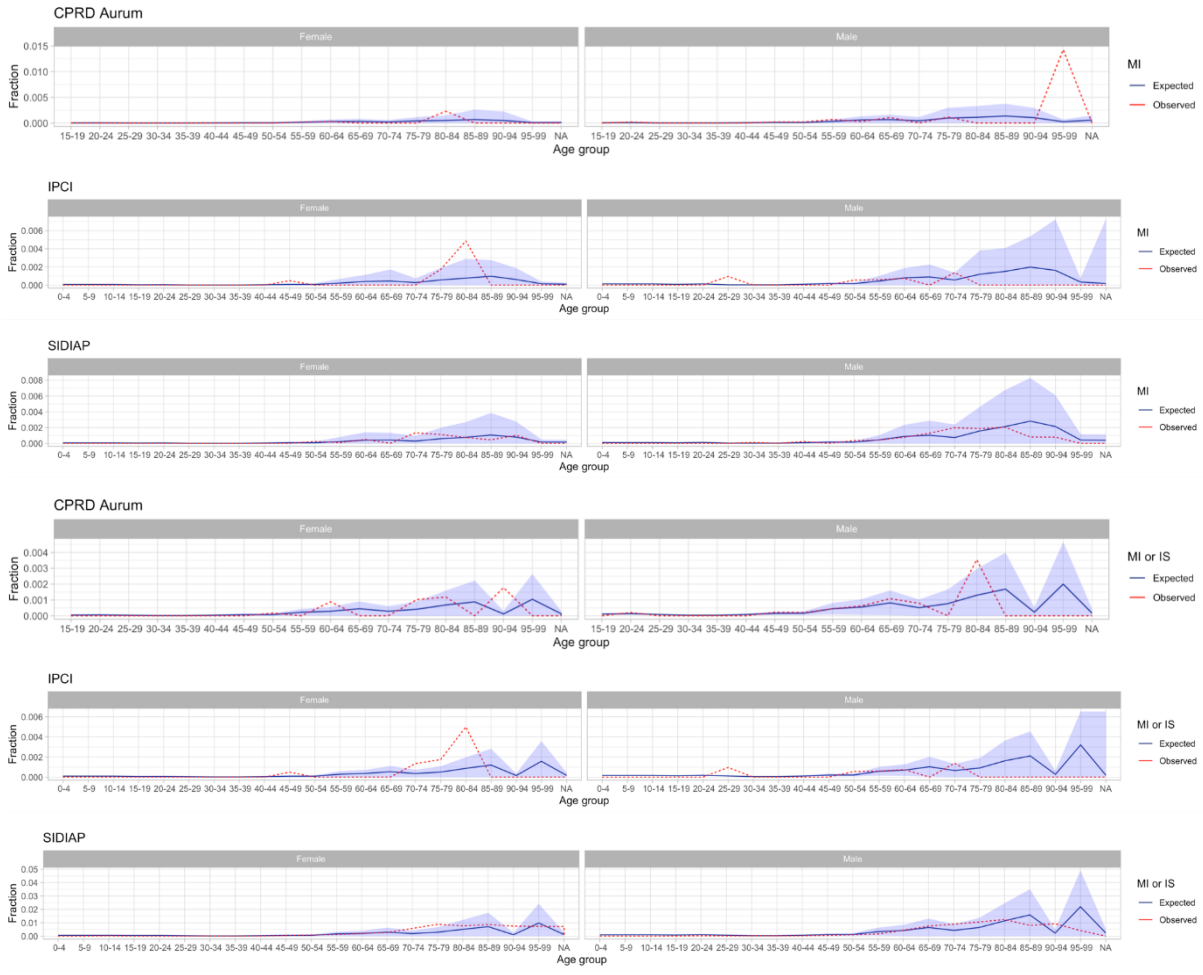
Table 5 - Coefficients in the parsimonious models for the prediction of 30-day MI and ATE

Predictor variable	MI	ATE
Intercept	-9.895531	-9.4757444
Sex, male	0.63659567	0.5320347
Autoimmune disease	0.16939754	0
Atrial fibrillation	0.25251097	0
Asthma	-0.47338496	0
Thrombophilia	0	0
Antiphospholipid syndrome	0	0
Malignant neoplastic disease	0.10887418	0
Diabetes mellitus	0.09328033	0
Obesity	-0.10525648	-0.1205930
Heart disease	0.27617190	0.5585278
Hypertensive disorder	0.45612847	0.5352648
Renal impairment	0.90337447	0.3788491
Chronic obstructive lung disease	0.55413691	0.1822488
Dementia	-1.82692282	-1.2350314
Antiinflammatory	0	0.3299078
Coxib	0.42325195	0
Corticosteroids	-0.54812656	-0.2752519
Antithrombotic agent	0.27380506	0.1593101
Lipid modifying agent	-0.14569790	-0.3477578
Antineoplastic immunomodulating	0.37486376	0
Hormonal contraceptives	-0.66529372	-0.2168920
Tamoxifen	1.70161422	0
Sex hormones modulators	-0.48185699	-0.8268795
Age 15-19	-0.44796242	-0.2375540
Age 20-24	0	0
Age 25-29	-1.49634802	-0.4889666
Age 30-34	-1.54330329	-1.3333241
Age 35-39	-1.55243197	-1.3467702
Age 40-44	-0.35770893	-0.5709500
Age 45-49	0.20843515	0
Age 50-54	0.08118682	0
Age 55-59	1.05978743	0.9730636
Age 60-64	1.56810239	1.1106247
Age 65-69	1.53124271	1.4392733
Age 70-74	0.96055533	0.9116707
Age 75-79	1.54033310	1.2310594
Age 80-84	1.66850135	1.7144092

Age 85-89	1.90076008	1.9983900
Age 90-94	1.58779893	0
Age 95-99	0	2.2260945
Age 100-104	0	0
Age 105-109	0	0

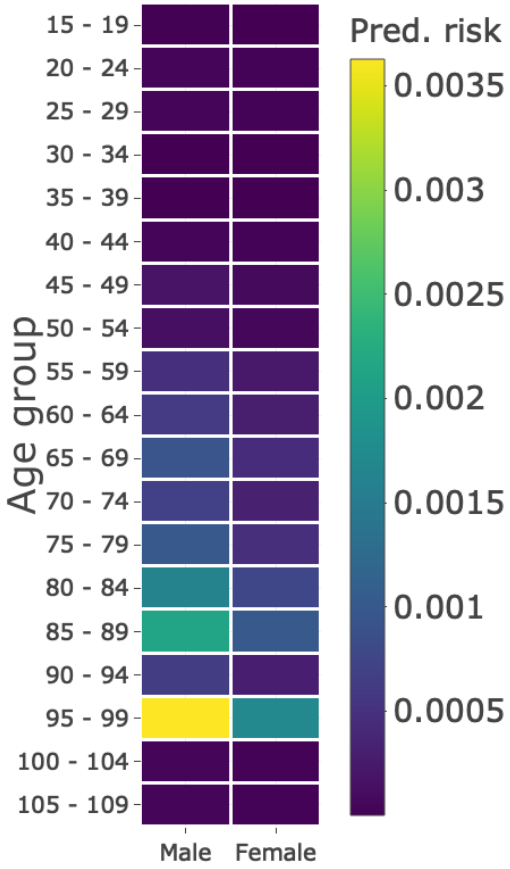
External validation findings were in line with those seen in the internal validation, and showed good discrimination for ATE with baseline models in COVID-19 patients aged <65 at the time of diagnosis/PCR (AUC 0.76 in NL, 0.83 in ES), but poorer in older patients (AUC 0.42 in NL, 0.56 in ES). The addition of pre-specified comorbidities/medicines use (parsimonious models) improved performance to a certain degree in NL data (AUC increased from 0.42 to 0.69) in line with UK findings, but not in ES (AUC 0.56 in baseline vs 0.55 in parsimonious models). See **Figure 9**. Calibration plots are depicted in **Figure 12**. The plotting of observed vs expected risks of MI and ATE stratified by age and sex confirmed good calibration for participants aged <65, but poorer performance in older patients.

Figure 12 - Age and sex-specific observed (red dash) vs expected (blue solid) risk of MI and ATE



Finally, **Figure 13** depicts a heatmap of predicted risk of 30-day ATE events amongst COVID-19 patients according to age and sex, based on the proposed baseline models. Like previously for VTE, these models should be used with caution, specially for patients aged 65 or older.

Figure 13 - Predicted risk of 30-day ATE amongst COVID-19 patients based on age and sex



14 Conclusions

14.1 Cumulative incidence of VTE/ATE

This is the largest cohort to date to study the risk and determinants of VTE and ATE amongst COVID-19 sufferers, including more than 1.1 million patients diagnosed with COVID-19 and/or tested + in a PCR for SARS-CoV-2 from 5 European countries. Data were obtained from primary care records from ES, NL, and the UK, and outpatient records from DE and FR. This is to our knowledge the first report on the risk/s of VTE/ATE amongst COVID-19 patients following them from an outpatient diagnosis. This is also by far the largest population analysed, with previous studies ranging from 12 to twelve thousand patients (1)(22). By leveraging information from the general (COVID-free) population and from outpatient and inpatient data, our analyses provide a more accurate picture of the natural history of thrombosis and coagulopathy in COVID-19, therefore minimising the risk of collider bias (2)(23).

Substantial heterogeneity in the background rates of both VTE and ATE events exists across databases (as described in previous documents). Despite the use of a common data model, the observed differences are reflective of the underlying variation in coding practice, healthcare settings included, and linkage availability across the participating databases.

Data from Spain (SIDIAP CMBD-AH) linked primary care to hospital records for a large cohort of almost half a million COVID-19 patients until June 2021, probably providing the most comprehensive real world data source on this topic in Europe to date. In this database, 30-day risk of VTE was 0.54%, mostly due to PE (cumulative incidence of 0.46%) and less so DVT (0.12%). Similarly, the 30-day risk of VTE post-diagnosis of or PCR+ for COVID-19 was 0.48%, mostly related to IS (0.39%) and less so due to MI events (cumulative incidence of 0.10%).

Most of the events were observed in the first 30 days of follow-up. For example, VTE risk increased only by about 20%, from 0.54% to 0.66%, from day 30 to 90, and ATE risk by almost 30%, from 0.48% to 0.61% in data from ES.

The occurrence of VTE and ATE prior to and following hospitalisation with COVID-19 was associated with worse prognosis for patients. Among those hospitalised, for example, the occurrence of VTE was associated with a hazard ratio of 1.63 (95% CI 1.39 to 1.90) for death, while ATE was associated with a hazard ratio of 1.93 (1.57 to 2.37).

Two recent systematic reviews have compiled information on numerous studies, but all of them focussed on hospitalized patients, with a substantial proportion (between 1 in 5 and one in 3) restricted to intensive care settings (1, 3)(22, 24). A systematic review of 102 previous studies by

Tan et al (1) has reported combined prevalence and incidence figures at an overall "frequency" of VTE of 14.7% [95%CI 12.1% to 17.6%] in hospitalized patients. Study-specific rates ranged from 0.0% to 69.2%, with meta-analytic 9.0% [6.9% to 11.4%] for patients hospitalized in general wards, and higher (meta-analytic estimate 23.2% [17.4% to 29.6%]) in ICU settings. A separate meta-analysis by S Nopp and colleagues (24) included 86 studies and reported a meta-analytic prevalence of 14.1% with ultrasound screening and 9.5% without, and again lower in non-ICU (7.9%) compared to ICU patients (22.7%). Our study showed much lower risk of VTE <1% amongst outpatient COVID-19 patients. This is likely due to a combination of younger age, a lower proportion of men, lower COVID-19 severity, and lack of screening in our cohorts compared to previous ones.

Frequency rates of ATE defined as a composite of IS and MI were reported in the study by Tan et al (22) at 3.0% [95%CI 2.7% to 3.4%], with study-specific estimates ranging from 0.0% to 8.9. As for VTE, our estimates of ATE were lower at about 0.5% in the 30-day period following an outpatient diagnosis or PCR+ for COVID-19. This difference is likely related to younger participants with a lower proportion of men and lower disease severity in our study compared to previous reports.

14.2 Associations between age, sex, and VTE/ATE

Men in our study had a higher risk of all the study outcomes, and risk also increased with increasing age in most analyses and databases. The association between age and VTE took an inverse U-shape in some data sources, with the increase in risk observed in ages 40-60, and a decreasing risk in older ages in most analyses. ATE had a different association with age, with small increases in risk in ages <70, but rapid increases in risk with older ages.

In general, risks of VTE and ATE can be expected to increase with age in the general population(4). Our finding of VTE peaking around 70 years of age in COVID-19 patients is likely explained by the substantial competing risk of mortality for those with COVID-19, which is much increased with older age.

In our study we found males to be at an increased risk of both ATE and, in all but one database, VTE. Although women are typically at higher risk of VTE before the menopause, being male has previously been associated with an increased risk of VTE in older ages. Similarly, men have also been seen to have an increased risk of ATE in previous literature (5)(26). Being male was also associated with an increased risk of mortality among those with COVID-19 in our study, consistent with previous research.

14.3 Associations between comorbidities, medicines use, and the risk of VTE/ATE

The study of comorbidities and medicines use in their association with VTE and ATE revealed interesting findings. Consistent increases in risk of VTE were seen in people with pre-existing asthma, cancer, obesity, and renal impairment. Similarly, we found a consistent association between prevalent systemic corticosteroid use and an increased risk of VTE. These are all well known risk factors for VTE in the general population, and these findings confirm a similar role amongst COVID-19 patients.

As for ATE, risk appeared consistently higher amongst people with a history of hypertension, diabetes, pervious heart disease, obesity and renal impairment. Smoking also appeared to confer a higher risk of ATE, as did the previous use of systemic glucocorticoids and, albeit less consistently, the previous treatment with anti-thrombotics.

Previous use of anti-thrombotics appeared associated with an increased risk of VTE in most, and with ATE in some of our analyses. Although inconsistent across databases, this association remained in some of our results despite adjustment for age and sex. This might be counterintuitive due to the known effects of these treatments to reduce the risk of thrombosis. However, this finding is likely the result of unresolved confounding by indication leading to paradoxical results that are opposite to intuition. This has been described by some as an example of Simpson's paradox (6)(27). Quoting from this paper by M Hernan et al "analytical errors may occur when the problem is stripped of its causal context and analyzed merely in statistical terms".

Although multivariable regression is a common strategy to minimise confounding, we pre-specified per protocol that we would only adjust our analyses for age and sex as we predicted that further adjustment would not only not be able to account for strong channelling bias but also create confusion. Similar multivariable models have been used in many COVID-19 studies previously, leading to counterintuitive findings like reductions in mortality associated with smoking(7)(28). The resulting models are prone to the so called "Table 2 fallacy" where ill-informed adjustment and potential overadjustment leads to wrong causal interpretations of the data (8)(29). Instead, we explicitly chose to adjust only for age and sex and to avoid causal interpretation of these analyses, which are merely descriptive in nature.

14.4 Prediction algorithms

Finally, we generated algorithms for the identification of patients at high risk of VTE or ATE in the 30 days after contracting COVID-19. Simple algorithms based only on a combination of age and sex had good performance for the estimation of risks of both VTE and ATE in patients aged <65

and in external validation exercises performed in all the contributing databases. Given the similar performance of complex algorithms compared to simple ones based on age and sex, the latter more parsimonious models were preferred for simplicity and ease of use.

Despite their relatively good performance overall, these models had lower discrimination for patients older than 65 years of age. It is possible that the drop in performance in older ages is related to bias in the derivation of our algorithms, which included cohorts enriched with younger patients. This likely resulted in insufficient statistical power to inform accurate models amongst older patients.

More research and potentially regulatory evaluations (as a potential medical device) are needed to further refine and improve these prediction algorithms before they can be recommended for use in clinical practice.

14.5 Strengths and Limitations

These analyses rely on routinely collected health care data from outpatient records. While this allowed us to include large study populations, the lack of linkage to hospital data is a limitation for all databases other than SIDIAP CMBDH-AH. Based on our previously reported analyses, this lack of hospital linkage can be expected to lead to incompleteness in the ascertainment of the study outcomes (here ATE and VTE) for the overall study populations. It also meant that outcomes for those hospitalised could only be described for SIDIAP CMBDH-AH. Despite this, the study of over 30,000 hospitalised patients from Spain constitutes the largest analysis of this topic to date, and the only one with sufficient detail on pre-admission stages of the natural history of COVID-19.

As explained above, analyses of so-called risk factors in COVID-19 are fraught with difficulties. Collider bias due, for example, to restricting analyses to people who have been diagnosed or hospitalised with COVID-19 can result in associations that do not exist in the general population or even reversing the sign of existing associations (22). Similarly, multivariable modelling involving the mutual adjustment for various factors of interest in the absence of a causal framework can lead to “Table 2 fallacy” and to paradoxical and counterintuitive findings. We pre-specified a simple adjustment strategy including only age and sex to clarify the descriptive nature of our data.

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

22. Tan BK, Mainbourg S, Friggeri A, Bertoletti L, Douplat M, Dargaud Y, et al. Arterial and venous thromboembolism in COVID-19: a study-level meta-analysis. Thorax. 2021;76:970-9.

23. Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nature Communications. 2020;11(1):5749.

24. Nopp S, Moik F, Jilma B, Pabinger I, Ay C. Risk of venous thromboembolism in patients

with COVID-19: A systematic review and meta-analysis. Research and Practice in Thrombosis and Haemostasis. 2020;4(7):1178-91.

25. Lowe GDO. Common risk factors for both arterial and venous thrombosis. British Journal of Haematology. 2008;140(5):488-95.
26. Millett ERC, Peters SAE, Woodward M. Sex differences in risk factors for myocardial infarction: cohort study of UK Biobank participants. BMJ. 2018;363:k4247.
27. Hernán MA, Clayton D, Keiding N. The Simpson's paradox unraveled. International Journal of Epidemiology. 2011;40(3):780-5.
28. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020;584(7821):430-6.
29. Westreich D, Greenland S. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. American Journal of Epidemiology. 2013;177(4):292-8.