Study Report

P3-C3-001

DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe

20/02/2025

Version 3.0

Contents

Study title	DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe
Study report version identifier	V3.0
Date of last version of report	20/02/2025
EU PAS number	EUPAS1000000254
Active substance	n/a
Medicinal product	n/a
Research question and objectives	This study aims to estimate the background incidence rates of selected vaccine adverse events of special interest, as well as to understand the demographic and clinical characteristics of patients with incident AESIs in Europe. 1. To estimate population level incidence rates of selected adverse events of special interest (AESIs) in the general population during 2010 and until the latest data availability, stratified by calendar year, month, sex, age groups, and data source. 2. To describe demographic and clinical characteristics of individuals with incident AESIs, and to compare the characteristics with individuals of similar age and sex but without the AESI.
Countries of study	UK, the Netherlands, Spain, Norway, Germany
Author(s)	Xintong Li, Edward Burn

TITLE

DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe

1. DESCRIPTION OF STUDY TEAM

Study team role(s)	Name(s)	Organisation(s)
Study Project Manager/Principal Investigator	Xintong Li, Edward Burn	University of Oxford
Data Scientist	Edward Burn	University of Oxford
Epidemiologist	Xintong Li Amy Lam	University of Oxford University of Oxford
Clinical Domain Expert	Daniel Prieto-Alhambra Albert Prats-Uribe George Corby James Bezer Abigail Robinson Ffion Samuels	University of Oxford University of Oxford University of Oxford University of Oxford University of Oxford University of Oxford
Data partner name*	**Data Partner member name(s)**	**Organisation(s)**
CPRD GOLD IPCI SIDIAP IQVIA DA Germany NLHR	Antonella Delmestri Mees Mosseveld Talita Duarte Salles James Brash Hedvig Nordeng	University of Oxford – CPRD data Erasmus MC – IPCI data IDIAPJGol – SIDIAP data IQVIA – Germany NLHR Norway – Uni Oslo

*Data partners' role is only to execute code at their data source, review and approve their results. These people do not have an investigator role. Data analysts/programmers do not have an investigator role and thus declaration of interests (DOI) for these people is not needed.

2. DATA SOURCES

Table 1. Description of data sources.

Country	Name of Database	Health Care setting	Type of Data	Number of individuals	Calendar period covered by each data source
Germany	IQVIA DA Germany	Primary care including specialists	EHR	5.25 million	1992-01-01 to 2023-09-30
The Netherlands	Integrated Primary Care Information (IPCI)	Primary care	EHR	1.24 million	2006-01-01 to 2023-12-31
Norway	Norwegian Linked Health Registry data (NLHR)	Primary care, including specialists, and secondary care	Registries	6.89 million	2008-01-01 to 2023-12-31
Spain	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP)	Primary care + linkage to hospital data	EHR	5.94 million	2006-01-01 to 2023-06-30
UK	Clinical Practice Research Datalink (CPRD) GOLD	Primary care	EHR	2.96 million	1987-09-09 to 2024-06-15

*Number of individuals were estimated by number of patients under observation as on 1st January 2023

3. ABSTRACT

Title

DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe

Rationale and background

Vaccines are approved for immunisation against various infectious diseases, with an increasing number based on recent novel platforms such as mRNA technology. Safety information for these new platforms was limited to pre-licensure clinical trials until the COVID-19 pandemic. The pandemic highlighted the need for timely post-authorisation vaccine safety surveillance for new vaccines and continuous monitoring throughout the lifecycle for established vaccines. Rapid regulatory responses to vaccine safety concerns are crucial for maintaining public confidence. Background incidence rates of adverse events of special interest (AESIs) can support these responses, with use in observed-to-expected analyses as an essential initial step in the continuum of signal evaluation.

The 2020 EMA-funded ACCESS project aimed to estimate the background rates of AESIs for monitoring COVID-19 vaccines. Since then, several publications have contributed to the global knowledge on background incidence rates, but regular updates are needed to remain prepared for new safety concerns. Granularity in estimates, particularly regarding risk groups and factors like seasonality, is important. Background rates vary across age groups, sex, regions, and data sources, influenced by different clinical coding systems and healthcare practices. Understanding patient demographics and clinical characteristics aids in evaluating potential safety signals. While some AESIs are specific to certain vaccines, others like Guillain-Barre syndrome are associated with various vaccines. This study aimed to expand previous research on background rates of selected AESIs to support future safety monitoring endeavours for both approved and newly developed vaccines, including as part of the EMA/ECDC Vaccine Monitoring Platform.

Research question and objectives

This study aimed to estimate the background incidence rates of selected vaccine adverse events of special interest, as well as understand the demographic and clinical characteristics of patients affected in Europe.

The list of AESIs was developed based on knowledge of most representative AESIs for a variety of vaccine safety signals (including for COVID-19 vaccines) and consultation with experts from EMA and EMA's Pharmacovigilance Risk Assessment Committee (PRAC).

Main objectives

1. To estimate population level incidence rates of selected adverse events of special interest (AESIs) in the general population during 2010 and until the latest data availability, stratified by calendar year, month, sex, age groups, and data source.
2. To estimate age and sex standardised incidence rates (to the European population) of selected adverse events of special interest (AESIs) in the general population during 2010 and until the latest data availability, stratified by calendar year.

Secondary objective

1. To describe demographic and clinical characteristics of individuals with incident AESIs and comparing the characteristics with individuals of similar age and sex but without the AESI.

Study design

Population-level cohort study.

Population:

All individuals observed in one of the participating data sources during the study period were eligible for inclusion. We required individuals to have at least 365 days of data availability before entering the cohort. The index date of cohort entry was study start or the date that individual fulfil the data availability and outcome 'clean window' requirement.

Variable:

The outcomes of this study were a pre-defined list of adverse events of special interest.

Data source:

1. Clinical Practice Research Datalink (CPRD) GOLD [UK]
2. Integrated Primary Care Information (IPCI) [The Netherlands]
3. Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP) [Spain]
4. Norwegian Linked Health Registry data (NLHR) [Norway]
5. IQVIA Disease Analyzer (DA) Germany [Germany]

Analysis:

Firstly, we developed the phenotypes for the study outcomes following the Standard Operating Procedure under a dynamic workflow between the study team and the EMA.

For Objective 1, incidence rates per 100,000 person-years were estimated along with their corresponding 95% confidence intervals. Analyses were stratified by calendar month, year, age group and sex within each database. Incidence rates were not estimated if there were less than 5 events in a given stratum. Incidence rates were also standardised to the European population for Objective 2.

For Objective 3, we used a large-scale characterisation (a summary at different time windows of all the conditions and drugs of the population) to describe the characteristics of each incident AESI cohort. For each incident AESI cohort, we summarised the demographics and clinical characteristics of individuals with incident outcomes using the large-scale characterisation. We constructed age-sex matched cohorts from the general population for each AESI cohort to contextualise the characteristics of the incident AESI cohort using standardised mean difference (SMD).

Results:

Among immune-mediated diseases, autoimmune thyroiditis and Bell's palsy were the most common outcomes, while Kawasaki disease and Guillain-Barré syndrome were the rarest outcomes. Thrombocytopenia was the most common blood disorder, although estimates varied substantially across databases. Arrythmia, coronary heart disease, and heart failure were the most common disorders of the cardiovascular system, while thrombotic microangiopathy and single organ cutaneous vasculitis were the rarest. While neuritis was one of the most common disorders of the nervous system, optic neuritis was one of the rarest. Among coagulation disorders, non-haemorrhagic stroke was most common, while cerebral venous sinus thrombosis and disseminated intravascular coagulation were the rarest outcomes. Meanwhile for disorders of other systems, tinnitus and hearing loss consistently had the highest incidence rates.

Most outcomes had higher incidence for older age groups, although Kawasaki, type 1 diabetes, and seizures were more common in the youngest age groups, autoimmune thyroiditis, multiple sclerosis, optic neuritis, and tinnitus peaked in middle age, while immune thrombocytopenia, thrombotic microangiopathy, and IgA vasculitis were more common in the youngest and oldest age groups. In general differences were less pronounced when stratifying by sex, although outcomes such as autoimmune thyroiditis and optic neuritis were common among females while outcomes such as Guillain-Barré syndrome, acute kidney injury, acute liver injury, and rhabdomyolysis were more frequently seen for males. Many outcomes had notable time trends in incidence rates, however these trends were mostly inconsistent and differed by database. Where different definitions for an outcome were considered, the most pronounced differences in estimates were seen for neuritis, cardiomyopathy, and encephalitis.

When comparing the characteristics of those individuals with a given AESI to individuals with a similar age and sex, those with the event were generally seen to have more comorbidities and prior medication use.

Discussion:

We estimated background incidence rates for a wide range of AESIs in five European databases. As well as estimating rates for populations as a whole, rates were stratified by age, sex, and calendar time. We also provided detailed cohort characteristics among people with the various conditions and contextualised the results by comparing to matched cohorts from the general population. However, the background rates need to be interpreted with caution given heterogeneity across databases and underlying time trends seen for many of the outcomes.

For any new studies aiming at using background rates for an emerging signal evaluation, it will be important to first assess if the phenotypes are fully aligned with the outcomes to be assessed, run diagnostics in the databases, and tailor as needed (e.g., considering information from spontaneous case reports and clinical case definitions).

4. LIST OF ABBREVIATIONS

Abbreviation	Name
AESI	Adverse events of special interest
ACCESS	Vaccine Covid-19 monitoring readiness
CDM	Common Data Model
COVID-19	Coronavirus disease-2019
CPRD	Clinical Practice Research Datalink
DA	Disease Analyzer
EHR	Electronic Health Record
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
GBS	Guillain-Barre syndrome
GDPR	General Data Protection Regulation
GP	General practice
GVDN	Global Vaccine Data Network
ICD	International classification of disease
ICD-10-CM	the International Classification of Diseases, Tenth Revision, Clinical Modification
IPCI	Integrated Primary Care Information
NLHR	Norwegian Linked Health Registry data
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
RT-PCR	Reverse transcription polymerase chain reaction
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SIDIAP	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària
SMD	standardised mean difference
SNOMED	Systematized Nomenclature of Medicine
SPEAC	Safety Platform for Emergency vACcines

5. AMENDMENTS AND UPDATES

None.

6. MILESTONES

Study deliverable	Timelines (planned)	Timelines (actual)
Draft Study Protocol	May 2024	6 June 2024
Final Study Protocol	June 2024	8 July 2024
Creation of Analytical code	August / September 2024	September 2024
Execution of Analytical Code on the data	October 2024	December 2024
Draft Study Report	November 2024	18 December 2024
Final Study Report	December 2024	12 February 2025
Draft Manuscript (if agreed on)		
Final Manuscript (if agreed on)		

7. RATIONALE AND BACKGROUND

Vaccines are approved and used for immunisation against various vaccine-preventable infectious diseases, with an increasing number based on recent novel manufacturing platforms (such as mRNA technology), for which safety experience was limited to pre-licensure clinical trials until the recent COVID-19 pandemic. As emphasised by the pandemic, there is a public health need for timely post-authorisation vaccine safety surveillance for new vaccines, but also for continuous monitoring along the product lifecycle for established vaccines.

A rapid initial regulatory response to a vaccine safety concern is crucial for maintaining public confidence in vaccination programs. Background incidence rates of adverse events of special interest (AESIs) may support the rapid initial response to a vaccine safety concern.[1] Observed-to-expected analyses are essential for such a response, in order to inform further steps of signal strengthening and evaluation.[2–4].

In 2020, the EMA-funded Vaccine Covid-19 monitoring readiness (ACCESS, EUPAS37273) project aimed at estimating background rates of AESIs for monitoring COVID-19 vaccines.[5,6] Since then, several publications from other research groups have contributed to the global knowledge regarding background incidence rates of AESIs. [7–10]However, there remains a need for regular updates to support readiness in case of a new safety concern, regardless of the type of vaccine product. There is also a need for higher granularity of estimates, especially with regards to groups at risk (e.g., age, comorbidities), as well as other factors such as seasonality/circulation of specific virus strains considered as independent risk factors for some events (e.g., addressed by stratification by month).

Previous studies have shown that background rates vary across age groups and sex, and are often heterogeneous between regions and data sources.[7–10] The heterogeneity can come from different clinical coding systems, health care delivery system, clinical practice, or reflect the true differences between the source population. It is therefore important to use background rates generated from the same or similar data source rather than rates estimated from different setting or data sources in observed-to-expected analysis. Understanding the demographic and clinical characteristics of patients could provide useful information for evaluating potential safety signals in the future.[11] Further, there's also a lack of acknowledgement and recognition of limitations with the utilised phenotypes and data, which calls for a more thorough assessment.

While some AESIs are considered specific to given vaccines, vaccine platforms, or classes of vaccines, several, such as Guillain-Barre syndrome (GBS) and other immune-mediated or neurological outcomes are listed across a wide range of vaccines. Therefore, the current study aims to expand the scope of previous studies to not only AESIs for approved vaccines, evidence generated from this study can also be used to support further newly developed vaccines. This study will support vaccine safety monitoring endeavours as part of the EMA/ECDC Vaccine Monitoring Platform[12].

8. RESEARCH QUESTION AND OBJECTIVES

This study aimed to estimate the background incidence rates of selected vaccine adverse events of special interest, as well as understand the demographic and clinical characteristics of patients in Europe.

Main objectives:

1. To estimate population level incidence rates of selected adverse events of special interest (AESIs) in the general population during 2010 and until the latest data availability, stratified by calendar year, month, sex, age groups, and data source.
2. To estimate age and sex standardised incidence rates (to the European population) of selected adverse events of special interest (AESIs) in the general population during 2010 and until the latest data availability, stratified by calendar year.

Secondary objective:

3. To describe demographic and clinical characteristics of individuals with incident AESIs and comparing the characteristics with individuals of similar age and sex but without the AESI.

9. RESEARCH METHODS

9.1 Study type and study design

Table 2. Description of potential study types and related study designs.

Study type	Study design	Study classification
Population-level descriptive epidemiology Patient-level characterisation	Population-level cohort	Complex *

*Note: This study is classified as Complex because of the high volume and the complexity of the phenotypes needed to be generated for the study.

This is a population-level retrospective, multi-database cohort study using electronic health record data from Europe. The incidence rates of AESIs were assessed using Population Level Disease Epidemiology.

9.2 Study setting and data sources

Database selection

The selection of databases for this study was performed based on data reliability and relevance for the proposed research question among those databases onboarded and available within DARWIN EU. The selected databases fulfil the criteria required for the availability of key information on exposures, outcomes, and covariates, while covering different settings and regions of Europe.

Table 3. Description of the selected data sources.

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active individuals*	Feasibility count of disease**	Data lock for the last update
Germany	IQVIA DA Germany	Database population representative of the general source population Good capture of some of the outcomes of interest	Primary care (GP and specialist)	EHR	5.25 million	deep venous thrombosis: 209100 GBS: 13300 Haemorrhagic stroke: 106100 Immune thrombocytopenia: 11700 Ischemic stroke: 528100 Acute myocardial infarction: 618000 narcolepsy: 20800 Pulmonary embolism: 272300 thrombocytopenia: 176700 Arrhythmia: 92200 Rheumatoid arthritis: 1695100 Type 1 diabetes mellitus: Tinnitus: Encephalitis: Immune thrombocytopenia: 11700 Pericarditis: 8500	2023-09-30
The Netherlands	Integrated Primary Care Information (IPCI)	Database population representative of the general source population. Good capture of some of the outcomes of interest	Primary care	EHR	1.24 million	deep venous thrombosis: 61500 GBS: 2400 Haemorrhagic stroke: 14600 Ischemic stroke: 95300 Acute myocardial infarction: 364300 Pulmonary embolism: 74200 thrombocytopenia: 6100 Arrhythmia: 11800 Rheumatoid arthritis: 127800	2023-12-31

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active individuals*	Feasibility count of disease**	Data lock for the last update
						Type 1 diabetes mellitus: 86000 Tinnitus: 122900 Encephalitis: 6300 Pericarditis: 5700	
Norway	Norwegian Linked Health Registry data (NLHR)	Linked national-wide population-level health registries. Almost the entire source population is captured. Good capture of some of the outcomes of interest	Primary care, including specialists, and secondary care	registries	6.89 million	deep venous thrombosis: 800 GBS: 100 Haemorrhagic stroke: 26100 Immune thrombocytopenia: 100 Ischemic stroke: 279300 Acute myocardial infarction: 540900 narcolepsy: 300 Pulmonary embolism: 267300 thrombocytopenia: 700 Arrhythmia: 21600 Rheumatoid arthritis: 1100800 Type 1 diabetes mellitus: 1431400 Tinnitus: 156700 Immune thrombocytopenia: 100 Pericarditis: 900 Autoimmune thyroiditis: 300	2023-12-31
Spain	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció	Database population representative of the general source population. Good capture of some	Primary care database + linkage to hospital	EHR	5.94 million	CVST: 700 deep venous thrombosis: 106100 GBS: 11600 Haemorrhagic stroke: 101500 Immune thrombocytopenia: 20200 Ischemic stroke: 448400 Acute myocardial infarction: 420700	2023-06-30

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active individuals*	Feasibility count of disease**	Data lock for the last update
	Primària (SIDIAP)	of the outcomes of interest	data			narcolepsy: 2400 Pulmonary embolism: 120400 thrombocytopenia: 202200 Arrhythmia: 56200 Rheumatoid arthritis: 111400 Type 1 diabetes mellitus: 200 Tinnitus: 238600 Immune thrombocytopenia: 20200 Pericarditis: 1500 Autoimmune thyroiditis: 25500	

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active individuals*	Feasibility count of disease**	Data lock for the last update
UK	Clinical Practice Research Datalink (CPRD) GOLD	Data with denominator populations representative of the general source population provide the most relevant and valid background rates.	Primary care	EHR	2.98 million	CVST: 1300 deep venous thrombosis: 183700 GBS: 3500 Haemorrhagic stroke: 23100 Immune thrombocytopenia: 11200 Ischemic stroke: 48400 Acute myocardial infarction: 295000 narcolepsy: 3600 Pulmonary embolism: 114100 thrombocytopenia: 52600 Arrhythmia: 38800 Rheumatoid arthritis: 218700 Type 1 diabetes mellitus: 106800 Tinnitus: 379600 Encephalitis: 900 Immune thrombocytopenia: 10700 Pericarditis: 100	2024-06-15

**Number of active individuals are estimated by number of patients under observation as on 1st January 2023 from the DARWIN Portal.*

*** Feasibility counts were estimated from record counts of concepts from the DARWIN Portal at study protocol stage.*

1) IQVIA Disease Analyser (DA) Germany [IQVIA DA Germany], Germany (IQVIA)

IQVIA Disease Analyzer (DA) Germany is a database of de-identified electronic medical records from specialized and general primary practices in Germany since 1992. [13] This dataset encompasses approximately 3% of all outpatient practices within Germany, ensuring a substantial representation of the national healthcare landscape. The sampling methods used for practice selection, taking into account physician's demographics, specialty focus, community size category and federal state location, was instrumental in constructing a database that accurately mirrors the diverse spectrum of healthcare providers in the country. Consequently, data within IQVIA DA Germany database has been demonstrated to be representative of general and specialised practices throughout Germany.

The database contains demographics records, basic medical data, disease diagnosis according to International Classification of Diseases, 10th revision (ICD-10), and prescription records. While the database partly records information on deaths and procedures, it currently does not support linkage with external data sources. Routine updates are conducted at regular intervals. The quality of data is assessed based on several criteria including completeness of information and correctness (e.g. linkage between diagnosis and prescriptions).

2) Integrated Primary Care Information Project [IPCI], The Netherlands (Erasmus MC)

The Integrated Primary Care Information (IPCI) database is a longitudinal observational database containing routinely collected data extracted from computer-based patient records of a selected group of general practitioners across the Netherlands. [14] IPCI was started in 1992 by the department of Medical Informatics of the Erasmus University Medical Center in Rotterdam. The current database includes patient records from 2006 on, when the size of the database started to increase significantly. The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex. Although the geographical spread is limited, GP practices are located in urban and non-urban areas.

Patient-level data includes demographic information, patient's complaints and symptoms, diagnoses, laboratory test results, lifestyle factors and correspondence with secondary care, such as referral and discharge letters.

3) Norwegian Linked Health Registry data [NLHR], Norway

Norway has a universal public health care system consisting of primary and specialist health care services covering a population of approximately 5.4 million inhabitants. Many population-based health registries were established in the 1960s with use of unique personal identifiers facilitating linkage between registries. Data in these health registries are used for health analysis, health statistics, improving the quality of healthcare, research, administration and emergency preparedness. We harmonized data from the following registries: the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Surveillance System for Communicable Diseases (MSIS), the Norwegian Immunisation Registry (SYSVAK), the National Death Registry, and the National Registry (NR). Linkage between the registries was facilitated using project-specific person ID generated from unique personal identification assigned at birth or immigration for all legal residents in Norway.

4) Information System for Research in Primary Care [SIDIAP], Spain (IDIAP Jordi Gol)

The Information System for Research in Primary Care (SIDIAP) is a dynamic database of pseudo-anonymized electronic health records of the primary care patient population in Catalonia, Spain. [15] It contains data of approximately 80% of the Catalan population registered in over 280 primary care practices throughout Catalonia since 2005.

The database contains data recorded in primary care centres on a daily basis. Additionally, it integrates data from external sources including biomarkers data from laboratories and records of drug prescription and dispensation. The dataset covers demographics, all-cause mortality, disease diagnoses classified under the ICD-10, prescription and dispensation records of drugs, results of laboratory tests, socio-economic indicators, vaccination records, lifestyle information, parent–child linkage and various clinical parameters. Additional data from other data sources such as hospital discharges, mental health centres or specific disease registries can be obtained through diverse linkages. The demographic composition within SIDIAP closely mirrors that of the broader Catalan population, encompassing a representative spectrum of geographic distribution, age, and sex proportions. The database is updated every 6 months.

5) Clinical Practice Research Datalink GOLD [CPRD GOLD], United Kingdom (University of Oxford)

The Clinical Practice Research Datalink (CPRD) GOLD is a database of anonymised electronic health records (EHR) from General Practitioner (GP) clinics in the UK that use the Vision® software system for their management. The source population encompasses 98% of the UK, registered with GPs responsible for non-emergency care and referrals. Participating GPs provide CPRD EHR for all registered patients who did not specifically request to opt out of data sharing. Covering 4.6% of the current UK population, GOLD includes 4.9% of contributing GP practices, providing comprehensive information within its defined source population. [16] GOLD contains data from all four UK constituent countries and the current regional distribution of its GP practices (among the 4.9% in the UK) is 5.7% in England, 55.6% in Scotland, 28.4% in Wales, and 10.2% in Northern Ireland (May 2022). GOLD data include patient's demographic, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.

9.3 Study period

From January 2010– until last available data, depending on the data sources. Two exceptions were made due to data availability/data quality reasons: for NLHR study start was 1st January 2018. For IQVIA DA Germany where follow-up was ended on 31st December 2022.

9.4 Follow-up

In the analysis of incidence rates, individuals began contributing person time on the latest of the following:

a) study start date,

b) date at which they have sufficient prior data availability (365 days), or

c) date on which they fulfil the 'clean window' (details in 8.6.2 Outcomes) criteria of a specific event.

Individuals were followed until the earliest date of the study events of interest, death, end of observation period in the database, or end of data availability of data source.

For acute and recurrent events, individuals were allowed to re-enter the same outcome cohort more than once, if they meet the inclusion criteria of data availability and 'clean window'. In contrast, time was censored after the first occurrence of a chronic event and the patient would not be allowed to re-enter the cohort.

9.5 Study population with in and exclusion criteria

General population:

The study population included all individuals observed in one of the participating data sources during the study period. We required individuals to have at least 365 days of data availability before entering the study.

9.6 Variables

9.6.1 Exposure /s

Not applicable.

9.6.2 Outcome/s

AESIs of interest:

The list was built on previously internationally recognized lists of AESIs by the Brighton Collaboration/ Safety Platform for Emergency vACcines (SPEAC) and curated by experts from EMA and EMA's Pharmacovigilance Risk Assessment Committee (PRAC), taking into account knowledge of most representative AESIs for a variety of vaccine safety signals (including for COVID-19 vaccines). Apart from AESIs included in previous studies, a broader list of conditions has been added. For example, conditions related to skin reactions are included. We excluded AESIs specific to one vaccine only and already well characterised (e.g., intussusception for rotavirus vaccines) or those which are very rare (e.g., multisystem inflammatory syndrome/MIS). The selected outcomes of interest are listed in Table 5.

For each study outcome, a clean window was applied to define incident outcomes. This was anytime prior in the patient history for chronic events and specific (shorter) clean windows for acute and recurrent outcomes (Table 5).

If the clean window was 90 days for a specific outcome, for example, the outcome event was considered incident if there was no record of the same outcome event during the preceding 90 days. An individual had the potential to contribute multiple outcome events if there was a gap of at least 90 days between each eligible event.

Table 5. Summary of event specific clean windows for outcomes.

	Outcome	Event specific clean window**	Type
Immune-mediated diseases	Guillain Barré syndrome	90	Acute
	Kawasaki disease	90	Acute
	Narcolepsy	NA	Chronic
	Immune Thrombocytopenia	90	Acute
	Type 1 diabetes	NA	Chronic

	Outcome	Event specific clean window**	Type
	Autoimmune thyroiditis	NA	Chronic
	Facial nerve palsy/Bells' palsy	90	Acute
Blood disorders	Thrombocytopenia	90	Acute
	Thrombotic thrombocytopenia syndrome	90	Acute
Disorders of the cardiovascular system	Coronary artery disease	NA	Chronic
	Heart failure	NA	Chronic
	Single organ cutaneous vasculitis	NA	Chronic
	Arrhythmia	NA	Chronic
	Thrombotic microangiopathy	90	Acute
	Cardiomyopathy	NA	Chronic
	Myocarditis	90	Acute
	Pericarditis	90	Acute
	Myocarditis/Pericarditis	90	Acute
Coagulation disorders	Cerebral venous thrombosis	90	Acute
	Deep vein thrombosis	90	Acute
	Disseminated intravascular coagulation	90	Acute
	Ischaemic stroke	90	Acute
	Haemorrhagic stroke	90	Acute
	Pulmonary embolism	90	Acute
Disorders of the nervous system	Epileptic convulsions/seizures	90	Acute
	Non-epileptic convulsions/seizures	90	Acute
	Febrile seizure	30	Recurrent
	Multiple sclerosis	NA	Chronic
	Acute Aseptic Meningitis	90	Acute
	Myelitis including transverse myelitis	90	Acute
	Encephalitis, which includes encephalomyelitis and ADEM	90	Acute
	Neuritis including optic neuritis	90	Acute

	Outcome	Event specific clean window**	Type
	Erythema multiforme	90	Acute
Disorders of the skin, bones and joints systems	Rheumatoid arthritis	NA	Chronic
	Drug reaction with eosinophilia and systemic symptoms	90	Acute
	Stevens-Johnson syndrome/ Toxic epidermal necrolysis	90	Acute
Disorders of other systems	Acute kidney injury	90	Acute
	Acute liver injury	90	Acute
	Anaphylaxis	30	Recurrent
	Pancreatitis (Acute)	90	Acute
	Rhabdomyolysis	90	Acute
	Sensorineural hearing loss	NA	Chronic
	Tinnitus	30	Recurrent
	Postmenopausal haemorrhage	30	Recurrent

**Event specific clean window: For each study outcome, a specific clean window was applied to define incident outcomes. NA identifies those outcomes that are chronic, and for these the washout window was any time in the patient history.

Phenotyping the AESIs

Firstly, we developed the phenotypes for the study outcomes following the Standard Operating Procedure under a dynamic workflow between the study team and the EMA [17] (**Figure 1**).

According to Hripcsak and Albers [18] "a phenotype is a specification of an observable, potentially changing state of an organism, […]. The term phenotype can be applied to patient characteristics inferred from electronic health record (EHR) data. […] . Phenotype algorithms – i.e., algorithms that identify or characterize phenotypes – may be generated by domain exerts and knowledge engineers, including recent research in knowledge engineering or through diverse forms of machine learning […] to generate novel representations of the data."

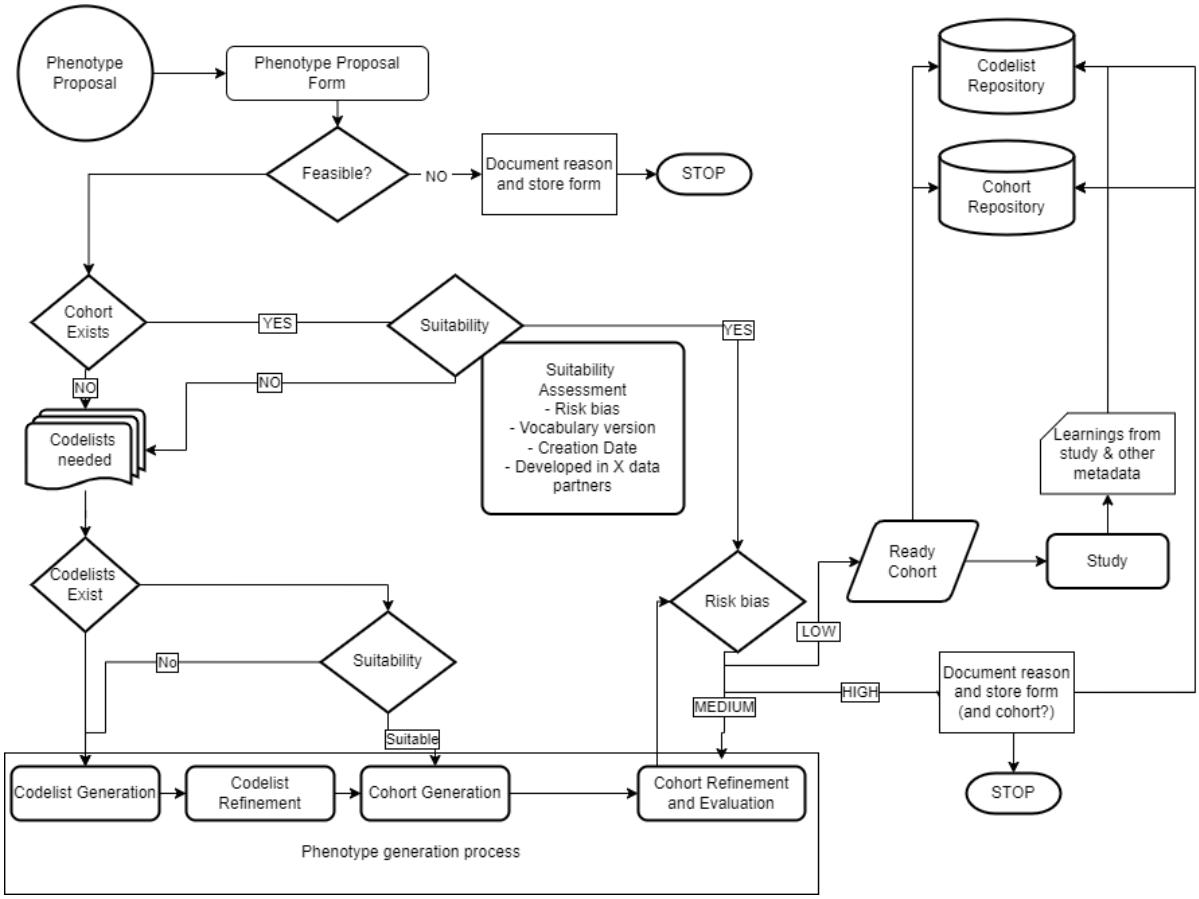

Figure 1. Summary diagram of the phenotyping process (from [17])

Firstly, the phenotype group of the DARWIN EU® Coordination Centre checked if a previous version of the *clinical description* exists in the DARWIN library. If not, a *clinical description* of the condition was created. The clinical description was intended to guide the phenotype development and to help evaluating the cohorts containing the phenotype, which should show characteristics of the disease of interest.

The *clinical description* was then reviewed internally by the senior clinical experts.

After the *clinical description* has been finalised, we checked whether a suitable cohort or phenotype already exists, that can be reused for this study. In case a compatible phenotype already exists, the next step is to decide whether it is suitable for the proposed use, or if it needs to be modified, and how. Depending on the answer to these questions, a phenotype can be reused as is or it can be modified or adapted for the proposed new use. All these decisions were documented.

In case where no compatible phenotype exists, a new one was generated. First, we undertook a *search* for potentially existing concept sets, and if available evaluated or modified for the proposed new use. If no concept list exists, a new *concept set* was then generated from scratch based on the submitted Phenotype Proposal Form, and similarly evaluated for use in DARWIN EU® studies and Data Partners.

Once the *concept sets* are available and deemed suitable, a *cohort* (or series of cohorts) was created based on that concept set, potentially including different flavours or modalities for different uses. Following this step, we ran the *diagnostics* over these cohorts. The results were evaluated and compared with the

characteristics from the clinical description, and further refinement of the phenotype was made if necessary. This is an iterative process until the phenotype is performing adequately. Finally, the phenotype was approved and stored for future.

For each purposed phenotype, the following information was documented:

- The proposed Logic (temporal, exclusion, etc.) and proposed Flavours (broad, narrow, etc.),
- The search strategy (keywords, domains)
- Concept sets/ Code list of each condition

In principle, three main flavours were considered: "Broad"- A broader definition of the outcome, which included codes that not very specific to the target phenotype, with higher sensitivity but lower specificity; "Narrow" – A narrower definition of outcome, with higher specificity at the cost of sensitivity; "Primary" – Primary outcome not lead by other known causes, or caused by medications or treatments.

9.6.3 Other covariates, including confounders, effect modifiers and other variables

Demographics:

Sex: Female, male.

Age

Age groups:

- 0– 4
- 5 –9
- 10-19
- 20 – 29
- 30 – 39
- 40 – 49
- 50 – 59
- 60 – 69
- 70 – 79
- 80 and over

A wider age group has also been used:

- 0-19
- 20 – 64
- 65 over

Health conditions:

We used large-scale characterisation to identify individuals´ history of the comorbidities using all available data prior to the index date (date of entering the general population cohort, or date of incident event). This is a data-driven method where all available data in the dataset will be utilized to provide a comprehensive

view of the potential comorbidities of an individual. Details of the large-scale characteristics is explained in Section 8.8 Analysis.

Medication use:

We identified the medication use of individual during the 183 days period before the index date. Similarly to conditions, the large-scale characterisation method was used.

9.7 Study size

The total population included in this study was approximately 76 million, estimated based on the "total number of patients" from all the included data sources as reported in the DARWIN portal. This does not reflect the actual number of individuals in each year during the study period, as the numbers of individuals in each data source is dynamic.

9.8 Data transformation

Not applicable.

9.9 Statistical methods

9.9.1 Main summary measures

In Objective 1, incidence rates with corresponding 95% confidence intervals were reported.

In Objective 2, patient-level demographical and clinical characteristics were summarised, and detailed in the section below.

9.9.2 Main statistical methods

Objectives 1: Estimating the incidence rates of AESIs

Incidence rates per 100,000 person-years was calculated as the number of incident cases divided by the total person-time at risk. The corresponding 95% confidence intervals was reported using exact Poisson model. A pre-defined clean window was applied to each AESI, an individual contributed to the person-time after the clean window of the previous event been fulfilled, except if the event is chronic.

The incidence rates were calculated using the "IncidencePrevalence" R package, developed by DARWIN EU.[19]

Stratification/subgroup

- By database
- By calendar year
- By time: pre-COVID-19 (1st January 2017 to 31st December 2019) and post-COVID-19 (1st January 2022 to end of follow-up).
- By age group (0– 4, 5 –9, 10-19, 20 – 29, 30 – 39, 40 – 49, 50 – 59, 60 – 69, 70 – 79, 80 and over). A wider age group has also been used (0-19, 20 – 64, and 65 over)

- By sex
- By age group and sex

Incidence rates were not estimated if there were less than 5 events in a given stratum.

We did not include the during COVID-19 to avoid the potential impact of restricted health care access, as well as impact from impact from COVID-19 infection and vaccinations especially those were linked to infection or listed in the Summary of Product Characteristic.

These incidence rates were further standardized by age and sex to the European population by the direct method, using the same age groups.

Objective 2: Population-level characteristics

For each incident AESI cohort, we summarised the demographics and clinical characteristics of individuals with incident outcomes using the large-scale characterisation.

To contextualise the characteristics of the incident AESI cohort, we constructed a matched cohort from the general population for each AESI cohort. We had an exact match on age and sex and required the matched individual to be under observation on the index date (diagnosis date) of the AESI cohort. We then used standardised mean difference (SMD) to contextualise the characteristics between the matched cohort and the AESI cohort.

This was only a descriptive characteristics analysis. Therefore, the aim of the matching was to provide a better context for the incident AESI cohort, rather than to achieve conditional exchangeability.

9.9.3 Missing values

Not applicable

9.9.4 Sensitivity analysis

Where multiple cohort definitions were created for a given phenotype, we reported results for each definition as part of the main results.

9.9.5 Deviation from original protocol

In NLHR (Norway) data, we observed that the incidence rates between 2008 to 2018 were significantly lower than in the later period. This was due to secondary care data availability only from 2018 onwards. Therefore, in the study, we restricted the analysis to 2018 onwards in this database, in order to use the more complete version of the database. Meanwhile, in IQVIA DA Germany, we observed that in the last months of follow-up there was a dramatic drop in the denominator population. This is because for this dataset, observation end date for an individual is based on their last visit, which introduces an artificial censoring. To avoid this issue, we ended follow-up in IQVIA DA Germany database at the end of 2022.

As well as calculating incidence rates for the study period as a whole, we also estimated incidence rates in two constrained periods: pre-covid (1st January 2017 to 31st December 2019) and post-covid (1st January 2022 to end of follow-up). This first period allowed for a comparison with previous studies undertaken during the Covid-19 pandemic to estimate background rates. The latter period allows for estimates from a time period closest to the current day, which may be particularly useful where calendar time trends in incidence rates are observed.

Two outcomes were added during phenotyping: non-haemorrhagic stroke and hearing loss. Non-haemorrhagic stroke was included as a broader version of ischaemic stroke, which included conditions that

were not haemorrhagic but could be ischaemic. Hearing loss was included as a broader version of sensorineural hearing loss, as many data sources do not have the granular code of sensorineural hearing loss.

10. DATA MANAGEMENT

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

The analytic code for this study was written in R and used standardised analytics wherever possible. Each data partner executed the study code against their database containing patient-level data, and then returned the results (csv files) which only contained aggregated data. The results from each of the contributing data sites were then be combined in tables and figures for the study report and web application.

11. QUALITY CONTROL

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, data partners will have run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides numerous 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. 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.

Study specific quality control

Objective 1 of this study includes the phenotyping process, where we creates the phenotypes of the study outcomes. This followed the standard procedure with all decision reviewed and documented. Complex phenotype definitions were discussed with the EMA.

12. RESULTS

All results for each outcome and database are available in the Shiny app at: https://data-dev.darwin-eu.org/aesi/ The Shiny app layout is as follows:

- Database details
    - Database descriptions – text descriptions of the participating databases.

    - Snapshot – a data-driven summary or the participating databases.
- Background incidence rates (Objective 1)
    - Incidence – tables and plot tabs presenting estimated incidence rates.
- Patient characteristics (Objective 2)
    - Cohort code use – a summary of the codes that led to inclusion into a study cohort.
    - Cohort characteristics – a summary of patient demographics for each of the study cohorts.
    - Large scale characteristics – a summary of all conditions and medications recorded for individuals in each of the study cohorts.
    - Compare large scale characteristics – a comparison of large-scale characteristics among those from a study cohort and age and sex matched individuals from the general population.
    - Cohort overlap – a summary of the overlap between different study cohorts.

12.1 AESI cohorts

12.1.1 Immune-mediated diseases

Eight cohorts were created for immune-mediated diseases: autoimmune thyroiditis (broad), autoimmune thyroiditis (narrow), Bell's palsy, Guillain-Barré syndrome, immune thrombocytopenia, Kawasaki disease, narcolepsy, and Type 1 diabetes. Compared to the narrow definition of autoimmune thyroiditis, the broader definition included unspecific codes of thyroiditis and acute thyroiditis which might not be an autoimmune condition. Type 1 diabetes was defined as having a new diagnosis of type 1 diabetes (including ketoacidosis), and with no history of a prior type 1 diabetes diagnosis or type 2 diabetes or use of any antidiabetic drug any time before type 1 diabetes diagnosis and had no record of any non-insulin antidiabetic drug in the year after type 1 diabetes diagnosis.

The largest cohort was for autoimmune thyroiditis (broad) in IQVIA DA Germany with 104,869 individuals included, while the smallest cohort was for Kawasaki disease in NLHR with 352 individuals included. No records were found for autoimmune thyroiditis (narrow), Kawasaki disease, or narcolepsy in IPCI. Median age was lowest for Kawasaki disease (between 4 and 5 years across the databases) and highest for Guillain-Barré syndrome and immune thrombocytopenia (between 56 and 64 years and 56 and 63 years, respectively, across the databases). Autoimmune thyroiditis (broad) and autoimmune thyroiditis (narrow) were most skewed towards females, with around 80% of the patients being female, while Guillain-Barré syndrome was reported in a majority of male individuals (with 55% to 62%, across data sources) (see **Table 6**).

Table 6. Cohorts for immune-mediated diseases.

		Database				
		CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Autoimmune thyroiditis (broad)						
Number records	-	4,153	1,010	104,869	7,757	33,835
Number subjects	-	4,153	1,010	104,869	7,757	33,835
Age (Median [IQR])	-	44 [32 - 55]	47 [36 - 57]	47[33 - 59]	43 [30 - 56]	46 [35 - 58]
Sex (n (%))	Female	3,510 (84.52%)	790 (78.22%)	86,267 (82.26%)	6,378 (82.22%)	28,600 (84.53%)
	Male	643 (15.48%)	220 (21.78%)	18,531 (17.67%)	1,379 (17.78%)	5,235 (15.47%)
	Unknown	-	-	71 (0.07%)	-	-
Autoimmune thyroiditis (narrow)						
Number records	-	2,146	-	94,087	5,751	15,898
Number subjects	-	2,146	-	94,087	5,751	15,898
Age (Median [IQR])	-	43 [32 - 55]	-	47 [33 - 59]	42 [29 - 55]	47 [36 - 59]
Sex (n (%))	Female	1,852 (86.30%)	-	78,047 (82.95%)	4,831 (84.00%)	13,772 (86.63%)
	Male	294 (13.70%)	-	15,975 (16.98%)	920 (16.00%)	2,126 (13.37%)
	Unknown	-	-	65 (0.07%)	-	-
Bell's palsy						
Number records	-	19,443	5,213	21,978	15,446	50,836
Number subjects	-	19,443	5,213	21,978	15,446	50,836
Age (Median [IQR])	-	48 [32 - 63]	53 [36 - 67]	58 [42 - 72]	49 [32 - 65]	56 [39 - 73]
Sex (n (%))	Female	9,764 (50.22%)	2,629 (50.43%)	11,438 (52.04%)	7,613 (49.29%)	25,268 (49.70%)
	Male	9,679 (49.78%)	2,584 (49.57%)	10,525 (47.89%)	7,833 (50.71%)	25,568 (50.30%)
	Unknown	-	-	15 (0.07%)	-	-
Guillain-Barré syndrome						
Number records	-	1,239	726	4,138	1,618	3,671
Number subjects	-	1,130	466	2,264	948	2,650
Age (Median [IQR])	-	58 [42 - 69]	56 [41 - 68]	64 [51 - 75]	57 [37 - 68]	62 [45 - 74]
Sex (n (%))	Female	532 (42.94%)	314 (43.25%)	1,846 (44.61%)	713 (44.07%)	1,407 (38.33%)
	Male	707 (57.06%)	412 (56.75%)	2,288 (55.29%)	905 (55.93%)	2,264 (61.67%)
	Unknown	-	-	<5	-	-
Immune thrombocytopenia						
Number records	-	4,551	1,870	5,740	6,892	10,773
Number subjects	-	3,878	999	3,683	3,350	6,968
Age (Median [IQR])	-	56 [29 - 73]	56 [33 - 70]	60 [38 - 74]	55 [29 - 72]	63 [36 - 78]
Sex (n (%))	Female	2,327 (51.13%)	1,081 (57.81%)	3,043 (53.01%)	3,654 (53.02%)	5,860 (54.40%)
	Male	2,224 (48.87%)	789 (42.19%)	2,694 (46.93%)	3,238 (46.98%)	4,913 (45.60%)
	Unknown	-	-	<5	-	-
Kawasaki disease						
Number records	-	399	-	908	568	635
Number subjects	-	384	-	621	352	559
Age (Median [IQR])	-	4 [2 - 6]	-	5 [3 - 10]	5 [3 - 10]	4 [2 - 7]
Sex (n (%))	Female	166 (41.60%)	-	405 (44.60%)	202 (35.56%)	217 (34.17%)

		Database				
		CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
	Male	233 (58.40%)	-	503 (55.40%)	366 (64.44%)	418 (65.83%)
Narcolepsy						
Number records	-	571	-	2,346	928	885
Number subjects	-	571	-	2,346	928	885
Age (Median [IQR])	-	37 [24 - 53]	-	49 [31 - 62]	30 [20 - 46]	41 [27 - 57]
Sex (n (%))	Female	336 (58.84%)	-	1,163 (49.57%)	523 (56.36%)	402 (45.42%)
	Male	235 (41.16%)	-	1,183 (50.43%)	405 (43.64%)	483 (54.58%)
Type 1 diabetes						
Number records	-	6,003	1,340	13,130	4,718	2,359
Number subjects	-	6,003	1,340	13,130	4,718	2,359
Age (Median [IQR])	-	15 [9 - 27]	22 [11 - 39]	47 [22 - 65]	23 [11 - 55]	26 [13 - 39]
Sex (n (%))	Female	2,498 (41.61%)	561 (41.87%)	6,818 (51.93%)	2,102 (44.55%)	1,039 (44.04%)
	Male	3,505 (58.39%)	779 (58.13%)	6,302 (48.00%)	2,616 (55.45%)	1,320 (55.96%)
	Unknown	-	-	10 (0.08%)	-	-

12.1.2 Blood disorders

Three cohorts were created for blood disorders: thrombocytopenia, thrombotic thrombocytopenia syndrome (broad), and thrombotic thrombocytopenia syndrome (narrow). Thrombocytopenia was defined using either a diagnosis code or measurement of platelet count. Compared to the narrow definition of thrombotic thrombocytopenia syndrome, the broader definition also included thrombosis (such as “Cerebrovascular accident”, “Embolism and thrombosis of the renal vein”, and “Axillary vein thrombosis”) accompanied by thrombocytopenia observed in the ten days either side of thrombosis.

The largest cohort was for thrombocytopenia in SIDIAP with 614,521 individuals included. The smallest cohort was for thrombotic thrombocytopenia syndrome (narrow) in IPCI with 517 individuals included. Median age for the thrombocytopenia cohorts ranged from 64 to 70, with all cohorts majority male (ranging from 54% to 63%). Meanwhile, median age for thrombotic thrombocytopenia syndrome (broad) and thrombotic thrombocytopenia syndrome (narrow) was between 70 and 77 and all cohorts were majority male (ranging from 59% to 72%), see **Table 7**.

Table 7. Cohort counts for blood disorders.

Variable name	Variable level	CDM name				
		CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Thrombocytopenia						
Number records	-	750,349	47,507	382,676	29,495	1,810,411
Number subjects	-	327,486	29,934	167,437	20,626	614,521
Age (Median [IQR])	-	69 [55 - 79]	68 [52 - 80]	70 [56 - 79]	64 [41 - 75]	70 [56 - 80]
Sex (n (%))	Female	286,367 (38.16%)	19,469 (40.98%)	141,990 (37.10%)	13,583 (46.05%)	709,813 (39.21%)
	Male	463,982 (61.84%)	28,038 (59.02%)	240,426 (62.83%)	15,912 (53.95%)	1,100,598 (60.79%)

Variable name	Variable level	CDM name: CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
	Unknown	-	-	260 (0.07%)	-	-
Thrombotic thrombocytopenia syndrome (broad)						
Number records	-	2,753	803	6,664	2,047	27,309
Number subjects	-	2,690	715	4,627	1,767	24,704
Age (Median [IQR])	-	73 [64 - 82]	77 [68 - 84]	76 [66 - 82]	71 [60 - 78]	75 [64 - 83]
Sex (n (%))	Female	930 (33.78%)	236 (29.39%)	2,088 (31.33%)	837 (40.89%)	8,953 (32.78%)
	Male	1,823 (66.22%)	567 (70.61%)	4,574 (68.64%)	1,210 (59.11%)	18,356 (67.22%)
	Unknown	-	-	<5	-	-
Thrombotic thrombocytopenia syndrome (narrow)						
Number records	-	2,278	576	5,070	1,766	24,487
Number subjects	-	2,227	517	3,541	1,540	22,592
Age (Median [IQR])	-	73 [63 - 81]	76 [67 - 83]	75 [65 - 82]	70 [59 - 78]	74 [64 - 83]
Sex (n (%))	Female	782 (34.33%)	159 (27.60%)	1,541 (30.39%)	718 (40.66%)	7,967 (32.54%)
	Male	1,496 (65.67%)	417 (72.40%)	3,528 (69.59%)	1,048 (59.34%)	16,520 (67.46%)
	Unknown	-	-	<5	-	-

12.1.3 Disorders of the cardiovascular system

Fourteen cohorts were created for cardiovascular system: arrhythmia, cardiomyopathy (broad), cardiomyopathy (primary), coronary artery disease: myocardial infarction, coronary artery disease: all, coronary artery disease: angina, coronary artery disease: atherosclerosis, heart failure, IgA vasculitis, myocarditis, myocarditis or pericarditis, pericarditis (broad), pericarditis (primary), single organ cutaneous vasculitis, and thrombotic microangiopathy. Arrhythmia included terms of non-congenital arrythmia with unspecified cause. . Compared to the primary definition of cardiomyopathy, the broader definition also included codes on hypertrophic cardiomyopathy and secondary cardiomyopathy with known cause such as “dilated cardiomyopathy secondary to alcohol”. Coronary artery disease is a broad condition and therefore was separated into four cohorts: myocardial infraction, angina, atherosclerosis, and a composite of all these three. Meanwhile, compared to the primary definition of pericarditis the broad definition also included pericarditis codes with known causes such as “post-infarction pericarditis”, “chronic rheumatic pericarditis”, and “disorder of pericardium”. Single organ cutaneous vasculitis was a very rare outcome, and many of the included databases did not have this level of granularity in the source data. Only “hypersensitivity angiitis” and “nodular vasculitis” were recorded in the participating databases.

The largest cohort was for arrythmia in IQVIA DA Germany with 500,987 individuals included, while the smallest cohort was for single organ cutaneous vasculitis in IQVIA DA Germany with 228 individuals included. No records for IgA vasculitis, single organ cutaneous vasculitis, or thrombotic microangiopathy were seen for IPCI. Cohorts for heart failure had the oldest median age (ranging from 75 to 80 across databases), while IgA vasculitis had the youngest cohort, although average age ranged dramatically across databases (from 14 to 63) as was also the case for thrombotic microangiopathy (median age varying from 28 to 73 across databases). Most cohorts were majority male, although a number of the cohorts for single organ cutaneous vasculitis were majority female (ranging from 66% to 47%) while all cohorts for thrombotic microangiopathy were majority female (ranging from 55% to 61%), see **Table 8**.

Table 8. Cohort counts for disorders of the cardiovascular system.

Variable name	Variable level	CDM name: CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Arrhythmia						
Number records	-	216,519	63,556	500,987	308,494	603,181
Number subjects	-	216,519	63,556	500,987	308,489	603,181
Age (Median [IQR])	-	74 [64 - 82]	71 [61 - 80]	69 [55 - 78]	66 [50 - 76]	73 [60 - 82]
Sex (n (%))	Female	99,260 (45.84%)	31,117 (48.96%)	257,293 (51.36%)	156,680 (50.79%)	278,065 (46.10%)
	Male	117,259 (54.16%)	32,439 (51.04%)	243,240 (48.55%)	151,814 (49.21%)	325,116 (53.90%)
	Unknown	-	-	454 (0.09%)	-	-
Cardiomyopathy (broad)						
Number records	-	11,152	2,809	37,993	16,269	41,443
Number subjects	-	11,152	2,809	37,993	16,269	41,443
Age (Median [IQR])	-	61 [50 - 71]	63 [52 - 73]	68 [57 - 77]	67 [53 - 76]	72 [60 - 81]
Sex (n (%))	Female	4,701 (42.15%)	1,158 (41.22%)	14,848 (39.08%)	6,979 (42.90%)	15,359 (37.06%)
	Male	6,451 (57.85%)	1,651 (58.78%)	23,123 (60.86%)	9,290 (57.10%)	26,084 (62.94%)
	Unknown	-	-	22 (0.06%)	-	-
Cardiomyopathy (primary)						
Number records	-	8,223	2,809	35,102	17,932	21,060
Number subjects	-	8,223	2,809	35,102	17,932	21,060
Age (Median [IQR])	-	60 [49 - 71]	63 [52 - 73]	68 [57 - 77]	67 [53 - 76]	70 [60 - 79]
Sex (n (%))	Female	3,193 (38.83%)	1,158 (41.22%)	13,603 (38.75%)	7,646 (42.64%)	5,725 (27.18%)
	Male	5,030 (61.17%)	1,651 (58.78%)	21,482 (61.20%)	10,286 (57.36%)	15,335 (72.82%)
	Unknown	-	-	17 (0.05%)	-	-
Coronary artery disease: all						
Number records	-	124,217	51,010	270,284	181,674	169,158
Number subjects	-	124,217	51,010	270,284	181,673	169,158
Age (Median [IQR])	-	68 [59 - 77]	68 [59 - 77]	69 [59 - 78]	69 [59 - 78]	71 [61 - 81]
Sex (n (%))	Female	45,687 (36.78%)	20,541 (40.27%)	112,753 (41.72%)	73,092 (40.23%)	57,783 (34.16%)
	Male	78,530 (63.22%)	30,469 (59.73%)	157,314 (58.20%)	108,582 (59.77%)	111,375 (65.84%)
	Unknown	-	-	217 (0.08%)	-	-
Coronary artery disease: angina						
Number records	-	45,401	29,513	103,751	142,031	82,886
Number subjects	-	45,401	29,513	103,751	142,028	82,886
Age (Median [IQR])	-	69 [60 - 77]	69 [60 - 78]	67 [56 - 77]	69 [60 - 78]	72 [62 - 81]

Variable name	Variable level	CDM name				
		CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Sex (n (%))	Female	18,365 (40.45%)	13,291 (45.03%)	48,985 (47.21%)	56,949 (40.10%)	31,788 (38.35%)
	Male	27,036 (59.55%)	16,222 (54.97%)	54,690 (52.71%)	85,082 (59.90%)	51,098 (61.65%)
	Unknown	-	-	76 (0.07%)	-	-
Coronary artery disease: atherosclerosis						
Number records	-	25,351	5,338	179,650	115,251	111,144
Number subjects	-	25,351	5,338	179,650	115,250	111,144
Age (Median [IQR])	-	68 [60 - 75]	68 [60 - 75]	71 [61 - 78]	70 [61 - 77]	70 [60 - 79]
Sex (n (%))	Female	7,891 (31.13%)	1,748 (32.75%)	67,431 (37.53%)	34,502 (29.94%)	28,864 (25.97%)
	Male	17,460 (68.87%)	3,590 (67.25%)	112,075 (62.39%)	80,749 (70.06%)	82,280 (74.03%)
	Unknown	-	-	144 (0.08%)	-	-
Coronary artery disease: myocardial infarction						
Number records	-	79,994	23,327	64,160	79,390	88,283
Number subjects	-	79,994	23,327	64,160	79,390	88,283
Age (Median [IQR])	-	69 [58 - 79]	68 [58 - 77]	69 [58 - 78]	71 [61 - 79]	70 [58 - 80]
Sex (n (%))	Female	28,001 (35.00%)	7,855 (33.67%)	22,945 (35.76%)	26,653 (33.57%)	26,858 (30.42%)
	Male	51,993 (65.00%)	15,472 (66.33%)	41,180 (64.18%)	52,737 (66.43%)	61,425 (69.58%)
	Unknown	-	-	35 (0.05%)	-	-
Heart failure						
Number records	-	89,622	33,515	273,300	179,875	294,793
Number subjects	-	89,622	33,515	273,300	179,874	294,793
Age (Median [IQR])	-	78 [69 - 85]	80 [71 - 86]	75 [65 - 82]	75 [66 - 83]	80 [72 - 87]
Sex (n (%))	Female	40,184 (44.84%)	17,336 (51.73%)	136,324 (49.88%)	83,559 (46.45%)	152,416 (51.70%)
	Male	49,438 (55.16%)	16,179 (48.27%)	136,796 (50.05%)	96,316 (53.55%)	142,377 (48.30%)
	Unknown	-	-	180 (0.07%)	-	-
IgA vasculitis						
Number records	-	9,189	-	10,469	1,991	10,585
Number subjects	-	9,189	-	10,469	1,991	10,585
Age (Median [IQR])	-	40 [8 - 73]	-	63 [28 - 76]	14 [5 - 49]	58 [13 - 76]
Sex (n (%))	Female	5,018 (54.61%)	-	5,377 (51.36%)	992 (49.82%)	5,943 (56.15%)
	Male	4,171 (45.39%)	-	5,091 (48.63%)	999 (50.18%)	4,642 (43.85%)
	None	-	-	-	-	-
Myocarditis						
Number records	-	1,582	248	22,707	12,911	2,213
Number subjects	-	1,509	151	15,533	9,344	2,081

Variable name	Variable level	CDM name				
		CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Age (Median [IQR])	-	43 [29 - 58]	48 [33 - 62]	52 [36 - 63]	54 [35 - 69]	47 [33 - 65]
Sex (n (%))	Female	576 (36.41%)	96 (38.71%)	10,104 (44.50%)	4,110 (31.83%)	617 (27.88%)
	Male	1,006 (63.59%)	152 (61.29%)	12,587 (55.43%)	8,801 (68.17%)	1,596 (72.12%)
	Unknown	-	-	16 (0.07%)	-	-
Myocarditis or pericarditis						
Number records	-	8,855	2,307	33,144	23,086	28,653
Number subjects	-	8,268	1,655	23,147	17,016	25,681
Age (Median [IQR])	-	50 [35 - 65]	55 [41 - 66]	56 [41 - 69]	58 [40 - 72]	60 [42 - 75]
Sex (n (%))	Female	2,872 (32.43%)	791 (34.29%)	15,822 (47.74%)	8,558 (37.07%)	10,557 (36.84%)
	Male	5,983 (67.57%)	1,516 (65.71%)	17,302 (52.20%)	14,528 (62.93%)	18,096 (63.16%)
	Unknown	-	-	20 (0.06%)	-	-
Pericarditis (broad)						
Number records	-	7,717	2,070	17,348	16,820	34,354
Number subjects	-	7,209	1,518	12,387	13,038	30,157
Age (Median [IQR])	-	51 [36 - 66]	55 [42 - 66]	63 [50 - 75]	60 [44 - 72]	61 [43 - 76]
Sex (n (%))	Female	2,442 (31.64%)	699 (33.77%)	8,939 (51.53%)	6,890 (40.96%)	13,221 (38.48%)
	Male	5,275 (68.36%)	1,371 (66.23%)	8,404 (48.44%)	9,930 (59.04%)	21,133 (61.52%)
	Unknown	-	-	5 (0.03%)	-	-
Pericarditis (primary)						
Number records	-	7,386	2,070	10,743	14,055	26,856
Number subjects	-	6,911	1,518	7,942	11,033	24,072
Age (Median [IQR])	-	51 [36 - 66]	55 [42 - 66]	67 [54 - 77]	60 [43 - 72]	61 [43 - 75]
Sex (n (%))	Female	2,328 (31.52%)	699 (33.77%)	5,902 (54.94%)	5,632 (40.07%)	10,013 (37.28%)
	Male	5,058 (68.48%)	1,371 (66.23%)	4,837 (45.02%)	8,423 (59.93%)	16,843 (62.72%)
	Unknown	-	-	<5	-	-
Single organ cutaneous vasculitis						
Number records	-	64	-	228	238	1,532
Number subjects	-	64	-	228	238	1,532
Age (Median [IQR])	-	60 [46 - 72]	-	56 [40 - 72]	66 [54 - 75]	70 [56 - 79]
Sex (n (%))	Female	42 (65.62%)	-	144 (63.16%)	130 (54.62%)	715 (46.67%)
	Male	22 (34.38%)	-	84 (36.84%)	108 (45.38%)	817 (53.33%)
Thrombotic microangiopathy						
Number records	-	327	-	2,994	788	1,508
Number subjects	-	297	-	2,089	484	1,181
Age (Median [IQR])	-	28 [9 - 56]	-	73 [59 - 80]	47 [19 - 64]	53 [34 - 69]
Sex (n (%))	Female	189 (57.80%)	-	1,836 (61.32%)	444 (56.35%)	825 (54.71%)

Variable name	Variable level	CDM name: CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
	Male	138 (42.20%)	-	1,153 (38.51%)	344 (43.65%)	683 (45.29%)
	Unknown	-	-	5 (0.17%)	-	-

12.1.4 Disorders of the nervous system

Fourteen cohorts were created for nervous system disorders: aseptic meningitis, encephalitis (broad), encephalitis (narrow), encephalitis (primary), febrile seizure, multiple sclerosis, myelitis (narrow), myelitis (primary), neuritis (broad), neuritis (narrow), neuritis (primary), non-epileptic seizures, optic neuritis, and seizure.

The aseptic meningitis was a relatively broad definition, where we included meningitis caused by virus such as “viral meningitis” and “herpes zoster with meningitis” as the more specific code of “aseptic meningitis” was only identified in CPRD GOLD. Three different definitions were used for encephalitis. The primary definition only included codes without specific causes of encephalitis. Compared to the primary definition of encephalitis, the narrow definition also included codes of infectious encephalitis such as “herpes zoster encephalitis”, “herpetic meningoencephalitis”, “viral encephalitis”, and “toxoplasma encephalitis”. The broad definition of encephalitis then also included codes such as “toxic encephalopathy” and “leukoencephalopathy”. We defined three cohorts for myelitis: broad, narrow, and primary. Compared to the primary definition of myelitis, which only included primary encephalitis, the broad definition also included codes with known causes such as “subacute necrotizing myelitis” and “myelitis due to herpes simplex”. Four cohorts were created for neuritis: broad, narrow, primary, and optic neuritis. The primary definition only included neuritis without specific cause. In the narrow definition, we also included infectious neuritis such as “tuberculous neuritis“. The broad definition included a broad code of “peripheral nerve disease” and peripheral neuropathies such as brachial, ulnar, lumbosacral neuritis. The cohort counts were similar for the narrow and the primary definition. A cohort of optic neuritis was created as well. Seizure was a broad definition where any seizure was included. Non-epileptic seizures were defined as people with seizure but no history of epilepsy before the index date.

The largest cohort was for neuritis (broad) in IQVIA DA Germany with 379,031 individuals included. The smallest cohort was for Myelitis (primary) in IPCI with 73 individuals included. No records for Encephalitis (primary), Encephalitis (narrow), or Neuritis (primary), or Optic neuritis were seen for IPCI. The youngest cohort was febrile seizure (median age between 2 and 3 across databases), while the oldest cohort was neuritis (broad) where median age ranged from 61 to 68 across databases. Most cohorts were relatively evenly split by sex, although most cohorts for aseptic meningitis, optic neuritis, myelitis (primary), and myelitis (narrow) were majority female, see **Table 9**.

Table 9. Cohort counts for disorders of the nervous system.

Variable name	Variable level	CDM name: CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Aseptic meningitis						
Number records	-	2,300	444	1,610	1,101	3,360
Number subjects	-	2,254	372	1,317	966	3,260
Age (Median [IQR])	-	32 [24 - 42]	35 [24 - 53]	36 [13 - 54]	49 [36 - 60]	29 [7 - 47]

		CDM name				
Variable name	Variable level	CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Sex (n (%))	Female	1,384 (60.17%)	251 (56.53%)	864 (53.66%)	778 (70.66%)	1,363 (40.57%)
	Male	916 (39.83%)	193 (43.47%)	745 (46.27%)	323 (29.34%)	1,997 (59.43%)
	Unknown	-	-	<5	-	-
Encephalitis (broad)						
Number records	-	1,841	1,681	25,896	4,608	5,937
Number subjects	-	1,723	1,196	16,952	3,157	5,460
Age (Median [IQR])	-	54 [32 - 70]	49 [22 - 67]	66 [49 - 78]	52 [34 - 67]	65 [44 - 77]
Sex (n (%))	Female	943 (51.22%)	819 (48.72%)	14,924 (57.63%)	2,301 (49.93%)	2,734 (46.05%)
	Male	898 (48.78%)	862 (51.28%)	10,952 (42.29%)	2,307 (50.07%)	3,203 (53.95%)
	Unknown	-	-	20 (0.08%)	-	-
Encephalitis (narrow)						
Number records	-	1,379	-	14,761	3,733	3,364
Number subjects	-	1,306	-	10,182	2,558	2,969
Age (Median [IQR])	-	52 [31 - 69]	-	53 [38 - 65]	49 [30 - 65]	54 [25 - 71]
Sex (n (%))	Female	721 (52.28%)	-	8,604 (58.29%)	1,960 (52.50%)	1,456 (43.28%)
	Male	658 (47.72%)	-	6,147 (41.64%)	1,773 (47.50%)	1,908 (56.72%)
	Unknown	-	-	10 (0.07%)	-	-
Encephalitis (primary)						
Number records	-	308	-	9,703	2,638	301
Number subjects	-	283	-	5,965	1,806	290
Age (Median [IQR])	-	46 [24 - 63]	-	53 [39 - 64]	50 [32 - 66]	38 [5 - 59]
Sex (n (%))	Female	153 (49.68%)	-	5,907 (60.88%)	1,406 (53.30%)	130 (43.19%)
	Male	155 (50.32%)	-	3,791 (39.07%)	1,232 (46.70%)	171 (56.81%)
	Unknown	-	-	5 (0.05%)	-	-
Febrile seizure						
Number records	-	11,747	3,204	17,496	6,132	12,938
Number subjects	-	9,841	2,411	10,209	4,934	11,613
Age (Median [IQR])	-	2 [1 - 3]	2 [1 - 3]	3 [2 - 5]	2 [1 - 3]	2 [1 - 2]
Sex (n (%))	Female	5,110 (43.50%)	1,348 (42.07%)	7,643 (43.68%)	2,666 (43.48%)	5,623 (43.46%)
	Male	6,637 (56.50%)	1,856 (57.93%)	9,837 (56.22%)	3,466 (56.52%)	7,315 (56.54%)
	Unknown	-	-	16 (0.09%)	-	-
Multiple sclerosis						
Number records	-	7,528	1,978	23,295	6,159	6,680
Number subjects	-	7,528	1,978	23,295	6,159	6,680
Age (Median [IQR])	-	45 [35 - 55]	46 [34 - 58]	47 [36 - 57]	42 [32 - 53]	43 [34 - 53]
Sex (n (%))	Female	5,307 (70.50%)	1,381 (69.82%)	17,160 (73.66%)	4,078 (66.21%)	4,474 (66.98%)
	Male	2,221 (29.50%)	597 (30.18%)	6,115 (26.25%)	2,081 (33.79%)	2,206 (33.02%)
	Unknown	-	-	20 (0.09%)	-	-
Myelitis (narrow)						
Number records	-	1,233	164	9,153	3,432	2,013
Number subjects	-	1,025	73	5,651	2,127	1,595

		CDM name				
Variable name	Variable level	CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Age (Median [IQR])	-	45 [32 - 58]	56 [44 - 64]	53 [39 - 65]	49 [34 - 64]	48 [35 - 62]
Sex (n (%))	Female	754 (61.15%)	96 (58.54%)	5,472 (59.78%)	1,919 (55.91%)	1,129 (56.09%)
	Male	479 (38.85%)	68 (41.46%)	3,678 (40.18%)	1,513 (44.09%)	884 (43.91%)
	Unknown	-	-	-	-	-
Myelitis (primary)						
Number records	-	1,231	164	9,142	3,413	1,958
Number subjects	-	1,023	73	5,644	2,117	1,551
Age (Median [IQR])	-	45 [32 - 58]	56 [44 - 64]	53 [39 - 65]	49 [34 - 65]	48 [35 - 61]
Sex (n (%))	Female	754 (61.25%)	96 (58.54%)	5,468 (59.81%)	1,914 (56.08%)	1,105 (56.44%)
	Male	477 (38.75%)	68 (41.46%)	3,671 (40.16%)	1,499 (43.92%)	853 (43.56%)
	Unknown	-	-	-	-	-
Neuritis (broad)						
Number records	-	49,488	58,162	589,726	346,066	101,174
Number subjects	-	42,895	40,679	379,031	195,151	85,629
Age (Median [IQR])	-	62 [49 - 73]	65 [52 - 75]	66 [52 - 77]	61 [49 - 73]	68 [54 - 78]
Sex (n (%))	Female	23,616 (47.72%)	31,277 (53.78%)	328,979 (55.79%)	189,518 (54.76%)	53,290 (52.67%)
	Male	25,872 (52.28%)	26,885 (46.22%)	260,416 (44.16%)	156,548 (45.24%)	47,884 (47.33%)
	Unknown	-	-	331 (0.06%)	-	-
Neuritis (narrow)						
Number records	-	6,656	-	406,843	212,374	56,907
Number subjects	-	6,013	-	237,330	113,691	48,339
Age	-	52 [37 - 67]	-	71.00 [59.00 - 79.00]	65 [53 - 75]	63 [50 - 75]
Sex (n (%))	Female	3,799 (57.08%)	-	222,919 (54.79%)	109,125 (51.38%)	26,922 (47.31%)
	Male	2,857 (42.92%)	-	183,750 (45.16%)	103,249 (48.62%)	29,985 (52.69%)
	Unknown	-	-	174 (0.04%)	-	-
Neuritis (primary)						
Number records	-	6,537	-	406,843	212,374	56,903
Number subjects	-	5,900	-	237,330	113,691	48,335
Age (Median [IQR])	-	51 [37 - 67]	-	71 [59 - 79]	65 [53 - 75]	63 [50 - 75]
Sex (n (%))	Female	3,749 (57.35%)	-	222,919 (54.79%)	109,125 (51.38%)	26,918 (47.31%)
	Male	2,788 (42.65%)	-	183,750 (45.16%)	103,249 (48.62%)	29,985 (52.69%)
	Unknown	-	-	174 (0.04%)	-	-
Non-epileptic seizures						
Number records	-	66,501	5,927	47,337	41,543	38,710
Number subjects	-	56,154	4,801	36,337	35,862	36,216
Age (Median [IQR])	-	34 [9 - 59]	3 [1 - 16]	29.00 [3.00 - 66.00]	25 [3 - 55]	20 [2 - 65]
Sex (n (%))	Female	30,722 (46.20%)	2,640 (44.54%)	24,151 (51.02%)	19,695 (47.41%)	17,875 (46.18%)
	Male	35,779 (53.80%)	3,287 (55.46%)	23,155 (48.92%)	21,848 (52.59%)	20,835 (53.82%)
	Unknown	-	-	31 (0.07%)	-	-
Optic neuritis						
Number records	-	2,798	-	4,623	3,466	6,488

		CDM name				
Variable name	Variable level	CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Number subjects	-	2,509	-	3,142	2,186	5,481
Age (Median [IQR])	-	39 [30 - 50]	-	44 [33 - 56]	42 [30 - 54]	59 [42 - 73]
Sex (n (%))	Female	1,979 (70.73%)	-	3,156 (68.27%)	2,360 (68.09%)	3,498 (53.91%)
	Male	819 (29.27%)	-	1,466 (31.71%)	1,106 (31.91%)	2,990 (46.09%)
	Unknown	-	-	-	-	-
Seizure						
Number records	-	88,500	6,434	88,612	104,217	47,020
Number subjects	-	67,133	5,087	47,089	56,206	42,192
Age (Median [IQR])	-	36 [15 - 58]	3 [1 - 24]	46 [9 - 67]	34 [14 - 57]	32 [3 - 67]
Sex (n (%))	Female	41,081 (46.42%)	2,858 (44.42%)	44,135 (49.81%)	50,661 (48.61%)	21,496 (45.72%)
	Male	47,419 (53.58%)	3,576 (55.58%)	44,434 (50.14%)	53,556 (51.39%)	25,524 (54.28%)
	Unknown	-	-	43 (0.05%)	-	-

12.1.5 Coagulation disorders

Eight cohorts were created for coagulation disorders: cerebral venous sinus thrombosis, deep vein thrombosis, disseminated intravascular coagulation, haemorrhagic stroke, ischaemic stroke, non-haemorrhagic stroke, pulmonary embolism (broad), pulmonary embolism (primary). In addition to ischaemic stroke, we also included a non-haemorrhagic stroke cohort, which included codes such as “cerebral infarction” which does not specify if it is ischaemic (in other words, this is a broader definition of ischemic stroke where if not specified as a haemorrhagic then it stroke is assumed to be ischaemic). Compared to the narrow definition of pulmonary embolism, the broad definition also included codes of pulmonary embolism with known cause such as “postoperative pulmonary embolus”, and “septic pulmonary embolism”.

The largest cohort was for non-haemorrhagic stroke in IQVIA DA Germany with 179,289 individuals included. The smallest cohort was for Disseminated intravascular coagulation in CPRD Gold with 110 individuals included. No records for cerebral venous sinus thrombosis or disseminated intravascular coagulation were seen for IPCI. Most cohorts were evenly split between male and females, but all cerebral venous sinus thrombosis cohorts were majority female (53% to 65%). The cerebral venous sinus thrombosis cohorts were also the youngest (with median age ranging from 47 to 54), while the ischaemic stroke and non-haemorrhagic stroke were the oldest (ranging from 72 to 77), see **Table 10**.

Table 10. Cohort counts for coagulation disorders.

		CDM name				
Variable name	Variable level	CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Cerebral venous sinus thrombosis						
Number records	-	797	-	614	841	483
Number subjects	-	737	-	386	573	460
Age (Median [IQR])	-	47 [30 - 63]	-	54 [39 - 70]	50 [34 - 64]	52 [36 - 69]
Sex (n (%))	Female	453 (56.84%)	-	399 (64.98%)	477 (56.72%)	256 (53.00%)
	Male	344 (43.16%)	-	215 (35.02%)	364 (43.28%)	227 (47.00%)

		CDM name				
Variable name	Variable level	CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Deep vein thrombosis						
Number records	-	63,847	19,665	98,639	29,152	72,010
Number subjects	-	56,093	11,951	60,619	21,830	61,718
Age (Median [IQR])	-	65 [50 - 77]	63 [51 - 74]	68 [55 - 78]	65 [52 - 75]	69 [56 - 80]
Sex (n (%))	Female	31,790 (49.79%)	10,041 (51.06%)	53,807 (54.55%)	13,024 (44.68%)	34,564 (48.00%)
	Male	32,057 (50.21%)	9,624 (48.94%)	44,782 (45.40%)	16,128 (55.32%)	37,446 (52.00%)
	Unknown	-	-	50 (0.05%)	-	-
Disseminated intravascular coagulation						
Number records	-	112	-	368	858	3,187
Number subjects	-	110	-	289	833	3,147
Age (Median [IQR])	-	58 [36 - 73]	-	62 [46 - 76]	64 [47 - 74]	65 [49 - 77]
Sex (n (%))	Female	74 (66.07%)	-	183 (49.73%)	388 (45.22%)	1,341 (42.08%)
	Male	38 (33.93%)	-	185 (50.27%)	470 (54.78%)	1,846 (57.92%)
Haemorrhagic stroke						
Number records	-	16,247	7,779	32,176	19,225	37,861
Number subjects	-	15,514	3,765	19,210	15,984	34,125
Age (Median [IQR])	-	68 [55 - 80]	68 [58 - 77]	69 [57 - 79]	73 [60 - 81]	73 [59 - 82]
Sex (n (%))	Female	8,351 (51.40%)	4,077 (52.41%)	15,733 (48.90%)	9,047 (47.06%)	17,770 (46.93%)
	Male	7,896 (48.60%)	3,702 (47.59%)	16,422 (51.04%)	10,178 (52.94%)	20,091 (53.07%)
	Unknown	-	-	21 (0.07%)	-	-
Ischaemic stroke						
Number records	-	33,551	31,391	148,350	114,960	187,729
Number subjects	-	32,160	11,289	69,264	65,937	155,537
Age (Median [IQR])	-	74 [64 - 82]	72 [62 - 80]	74 [63 - 81]	73 [64 - 80]	77 [66 - 85]
Sex (n (%))	Female	15,454 (46.06%)	14,727 (46.91%)	67,591 (45.56%)	48,389 (42.09%)	89,500 (47.68%)
	Male	18,097 (53.94%)	16,664 (53.09%)	80,708 (54.40%)	66,571 (57.91%)	98,229 (52.32%)
	Unknown	-	-	51 (0.03%)	-	-
Non-haemorrhagic stroke						
Number records	-	91,001	94,559	393,032	434,278	226,760
Number subjects	-	83,994	39,721	179,289	131,480	168,924
Age (Median [IQR])	-	74 [64 - 83]	73 [63 - 81]	74 [64 - 81]	73 [64 - 80]	77 [67 - 85]
Sex (n (%))	Female	43,202 (47.47%)	45,903 (48.54%)	183,580 (46.71%)	182,181 (41.95%)	105,703 (46.61%)
	Male	47,799 (52.53%)	48,656 (51.46%)	209,267 (53.24%)	252,097 (58.05%)	121,057 (53.39%)
	Unknown	-	-	185 (0.05%)	-	-
Pulmonary embolism (narrow)						
Number records	-	47,917	19,427	116,781	196,034	51,184
Number subjects	-	43,970	11,691	60,991	60,976	45,192
Age (Median [IQR])	-	68 [55 - 78]	64 [51 - 74]	71 [59 - 80]	69 [57 - 77]	72 [60 - 82]
Sex (n (%))	Female	25,175 (52.54%)	10,453 (53.81%)	63,225 (54.14%)	82,685 (42.18%)	25,541 (49.90%)
	Male	22,742 (47.46%)	8,974 (46.19%)	53,505 (45.82%)	113,349 (57.82%)	25,643 (50.10%)
	Unknown	-	-	51 (0.04%)	-	-

Variable name	Variable level	CDM name: CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Pulmonary embolism (primary)						
Number records	-	47,824	19,427	116,781	196,034	50,923
Number subjects	-	43,889	11,691	60,991	60,976	44,944
Age (Median [IQR])	-	68 [55 - 78]	64 [51 - 74]	71 [59 - 80]	69 [57 - 77]	73 [60 - 82]
Sex (n (%))	Female	25,126 (52.54%)	10,453 (53.81%)	63,225 (54.14%)	82,685 (42.18%)	25,455 (49.99%)
	Male	22,698 (47.46%)	8,974 (46.19%)	53,505 (45.82%)	113,349 (57.82%)	25,468 (50.01%)
	Unknown	-	-	51 (0.04%)	-	-

12.1.6 Disorders of the skin, bones and joints systems

Four cohorts were created for skin, bones and joints system: drug reaction with eosinophilia and systemic symptoms, erythema multiforme, rheumatoid arthritis, and Stevens-Johnson syndrome/ toxic epidermal necrolysis.

The largest cohort was for rheumatoid arthritis in IQVIA DA Germany with 118,945 individuals included. The smallest cohort was drug reaction with eosinophilia and systemic symptoms in CPRD Gold with 45 individuals included. No records for erythema multiforme or Stevens-Johnson syndrome/ toxic epidermal necrolysis were seen for IPCI, while no records for drug reaction with eosinophilia and systemic symptoms were seen for IPCI, IQVIA DA Germany, and NLHR. All cohorts were majority female, except for Stevens-Johnson syndrome/ toxic epidermal necrolysis which was relatively evenly split between males and females. The erythema multiforme had the youngest average age (ranging from 26 to 38 years across databases), while median age of rheumatoid arthritis cohorts ranged between 60 and 63, see **Table 11**.

Table 11. Cohort counts for disorders of the skin, bones and joints systems.

Variable name	Variable level	CDM name: CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Drug reaction with eosinophilia and systemic symptoms						
Number records	-	46	-	-	-	52
Number subjects	-	45	-	-	-	52
Age (Median [IQR])	-	76 [67 - 83]	-	-	-	59 [43 - 75]
Sex (n (%))	Female	44 (95.65%)	-	-	-	32 (61.54%)
	Male	-	-	-	-	20 (38.46%)
Erythema multiforme						
Number records	-	5,288	-	3,960	1,048	6,124
Number subjects	-	4,944	-	3,535	885	5,796
Age	-	26 [7 - 49]	-	38.00 [11.00 - 59.00]	38 [15 - 59]	26 [5 - 53]
Sex (n (%))	Female	3,046 (57.60%)	-	2,294 (57.93%)	564 (53.82%)	3,285 (53.64%)
	Male	2,242 (42.40%)	-	1,666 (42.07%)	484 (46.18%)	2,839 (46.36%)
Rheumatoid arthritis						

		CDM name				
Variable name	Variable level	CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Number records	-	26,471	13,330	118,945	118,472	27,302
Number subjects	-	26,471	13,330	118,945	118,469	27,302
Age (Median [IQR])	-	61 [49 - 71]	61 [49 - 73]	60 [50 - 72]	63 [49 - 74]	63 [50 - 75]
Sex (n (%))	Female	17,739 (67.01%)	8,551 (64.15%)	80,770 (67.91%)	67,935 (57.34%)	19,234 (70.45%)
	Male	8,732 (32.99%)	4,779 (35.85%)	38,120 (32.05%)	50,537 (42.66%)	8,068 (29.55%)
	Unknown	-	-	55 (0.05%)	-	-
Stevens-Johnson syndrome/ toxic epidermal necrolysis						
Number records	-	474	-	291	347	502
Number subjects	-	437	-	258	287	431
Age (Median [IQR])	-	42 [22 - 63]	-	60 [42 - 75]	51 [29 - 67]	58 [35 - 74]
Sex (n (%))	Female	251 (52.95%)	-	165 (56.70%)	169 (48.70%)	250 (49.80%)
	Male	223 (47.05%)	-	125 (42.96%)	178 (51.30%)	252 (50.20%)

12.1.7 Disorders of other systems

Thirteen cohorts were created for others: acute kidney injury (broad), acute kidney injury (primary), acute liver injury, acute pancreatitis (broad), acute pancreatitis (narrow), anaphylaxis, hearing loss, postmenopausal haemorrhage (broad), postmenopausal haemorrhage (narrow), rhabdomyolysis (broad), rhabdomyolysis (narrow), sensorineural hearing loss, and tinnitus. Compared to the primary definition of acute kidney injury, the overall definition also included codes with known causes such as “acute renal failure due to obstruction” and “postoperative renal failure”, and broader code of “injury of kidney”. Compared to the narrow definition of acute pancreatitis, the broad definition also included codes such as “pseudocyst of pancreas”, “gallstone acute pancreatitis”, and “biliary acute pancreatitis”. The anaphylaxis cohort was defined by unspecific anaphylaxis code or anaphylaxis caused by drug or vaccine. The sensorineural hearing loss cohort is limited to diagnosis codes of sensorineural hearing loss. We further defined the hearing loss cohort where unspecific hearing loss was included. For the postmenopausal haemorrhage cohort, we restricted to women aged 45 or older. Compared to the narrow definition of postmenopausal haemorrhage, the broad definition also included codes such as “abnormal uterine bleeding”, “vaginal bleeding”, and “genitourinary tract hemorrhage”. The narrow definition of rhabdomyolysis included non-trauma rhabdomyolysis, whiles the broad definition also included codes such as “traumatic rhabdomyolysis” and “crush syndrome”.

The largest cohort was for hearing loss in IQVIA DA Germany with 444,341 individuals included. The smallest cohort was for acute liver injury in CPRD GOLD with 782 individuals included. No records for acute kidney injury (overall), acute kidney injury (primary), acute liver injury, rhabdomyolysis (broad), rhabdomyolysis (non-trauma), or sensorineural hearing loss were seen for IPCI. By definition, the postmenopausal haemorrhage (narrow) and postmenopausal haemorrhage (narrow) only included women. Meanwhile all acute liver injury cohorts were majority male (ranging from 55% to 64% across databases). The anaphylaxis cohorts were the youngest (with median age ranging from 39 to 50 across the databases) while the acute kidney injury (primary) cohorts were the oldest (average age from 71 to 79), see Table 12.

Table 12. Cohort counts for disorders of other systems.

Variable name	Variable level	CDM name: CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Acute kidney injury (broad)						
Number records	-	73,220	-	29,802	105,065	318,688
Number subjects	-	64,442	-	21,333	88,108	246,941
Age (Median [IQR])	-	74 [61 - 83]	-	71 [58 - 81]	76 [65 - 84]	79 [69 - 86]
Sex (n (%))	Female	36,113 (49.32%)	-	14,371 (48.22%)	43,109 (41.03%)	137,671 (43.20%)
	Male	37,107 (50.68%)	-	15,392 (51.65%)	61,956 (58.97%)	181,017 (56.80%)
	Unknown	-	-	39 (0.13%)	-	-
Acute kidney injury (primary)						
Number records	-	68,239	-	25,460	98,722	316,667
Number subjects	-	60,064	-	17,835	83,432	245,456
Age (Median [IQR])	-	74 [62 - 83]	-	73 [60 - 82]	76 [66 - 84]	79 [69 - 86]
Sex (n (%))	Female	33,740 (49.44%)	-	12,677 (49.79%)	40,683 (41.21%)	136,933 (43.24%)
	Male	34,499 (50.56%)	-	12,751 (50.08%)	58,039 (58.79%)	179,734 (56.76%)
	Unknown	-	-	32 (0.13%)	-	-
Acute liver injury						
Number records	-	802	-	18,278	11,275	31,854
Number subjects	-	782	-	11,045	8,538	25,692
Age (Median [IQR])	-	55 [44 - 66]	-	63 [53 - 73]	64 [53 - 73]	66 [55 - 75]
Sex (n (%))	Female	357 (44.51%)	-	7,870 (43.06%)	5,067 (44.94%)	11,473 (36.02%)
	Male	445 (55.49%)	-	10,404 (56.92%)	6,208 (55.06%)	20,381 (63.98%)
	Unknown	-	-	<5	-	-
Acute pancreatitis (broad)						
Number records	-	24,737	6,286	56,206	21,386	56,364
Number subjects	-	20,808	3,867	39,814	17,055	43,953
Age (Median [IQR])	-	57 [43 - 71]	61 [49 - 72]	60 [48 - 72]	60 [46 - 73]	65 [50 - 78]
Sex (n (%))	Female	11,910 (48.15%)	2,773 (44.11%)	27,824 (49.50%)	9,782 (45.74%)	26,295 (46.65%)
	Male	12,827 (51.85%)	3,513 (55.89%)	28,349 (50.44%)	11,604 (54.26%)	30,069 (53.35%)
	Unknown	-	-	33 (0.06%)	-	-
Acute pancreatitis (narrow)						
Number records	-	24,117	6,286	54,259	20,782	53,176
Number subjects	-	20,475	3,867	38,716	16,760	41,910
Age (Median [IQR])	-	57 [43 - 71]	61 [49 - 72]	59 [48 - 72]	60 [46 - 73]	65 [49 - 78]
Sex (n (%))	Female	11,655 (48.33%)	2,773 (44.11%)	26,801 (49.39%)	9,502 (45.72%)	24,812 (46.66%)
	Male	12,462 (51.67%)	3,513 (55.89%)	27,426 (50.55%)	11,280 (54.28%)	28,364 (53.34%)

Variable name	Variable level	CDM name: CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
	Unknown	-	-	32 (0.06%)	-	-
Anaphylaxis						
Number records	-	6,720	3,005	13,606	22,562	9,670
Number subjects	-	6,094	1,599	9,604	10,861	8,650
Age (Median [IQR])	-	39 [21 - 56]	50 [24 - 63]	47 [24 - 61]	44 [25 - 59]	42 [20 - 59]
Sex (n (%))	Female	3,938 (58.60%)	1,518 (50.52%)	8,048 (59.15%)	12,819 (56.82%)	5,019 (51.90%)
	Male	2,782 (41.40%)	1,487 (49.48%)	5,527 (40.62%)	9,743 (43.18%)	4,651 (48.10%)
	Unknown	-	-	31 (0.23%)	-	-
Hearing loss						
Number records	-	236,559	51,975	444,341	253,460	414,634
Number subjects	-	236,559	51,975	444,341	253,453	414,634
Age (Median [IQR])	-	64 [45 - 76]	69 [56 - 79]	61 [43 - 75]	65 [51 - 74]	63 [45 - 77]
Sex (n (%))	Female	119,297 (50.43%)	25,752 (49.55%)	234,416 (52.76%)	124,067 (48.95%)	216,962 (52.33%)
	Male	117,262 (49.57%)	26,223 (50.45%)	209,715 (47.20%)	129,393 (51.05%)	197,672 (47.67%)
	Unknown	-	-	210 (0.05%)	-	-
Postmenopausal haemorrhage (broad)						
Number records	-	131,781	26,482	74,189	83,292	72,661
Number subjects	-	106,764	19,260	57,325	56,552	67,056
Age (Median [IQR])	-	55 [51 - 64]	59 [54 - 68]	54 [50 - 62]	58 [53 - 68]	57 [52 - 69]
Sex (n (%))	Female	131,781 (100.00%)	26,482 (100.00%)	74,189 (100.00%)	83,292 (100.00%)	72,661 (100.00%)
Postmenopausal haemorrhage (narrow)						
Number records	-	62,438	26,482	33,331	72,536	64,662
Number subjects	-	54,223	19,260	27,176	49,051	60,126
Age (Median [IQR])	-	58 [54 - 66]	59 [54 - 68]	59 [55 - 67]	59 [54 - 70]	58 [53 - 70]
Sex (n (%))	Female	62,438 (100.00%)	26,482 (100.00%)	33,331 (100.00%)	72,536 (100.00%)	64,662 (100.00%)
Rhabdomyolysis (broad)						
Number records	-	1,979	-	-	-	14,670
Number subjects	-	1,917	-	-	-	14,058
Age (Median [IQR])	-	69 [42 - 82]	-	-	-	77 [56 - 85]
Sex (n (%))	Female	737 (37.24%)	-	-	-	6,291 (42.88%)
	Male	1,242 (62.76%)	-	-	-	8,379 (57.12%)
Rhabdomyolysis (narrow)						
Number records	-	1,732	-	-	-	14,670
Number subjects	-	1,674	-	-	-	14,058

Variable name	Variable level	CDM name: CPRD GOLD	IPCI	IQVIA DA Germany	NLHR	SIDIAP
Age (Median [IQR])	-	70 [41 - 82]	-	-	-	77 [56 - 85]
Sex (n (%))	Female	636 (36.72%)	-	-	-	6,291 (42.88%)
	Male	1,096 (63.28%)	-	-	-	8,379 (57.12%)
Sensorineural hearing loss						
Number records	-	44,323	-	186,872	216,083	42,841
Number subjects	-	44,323	-	186,872	216,079	42,841
Age (Median [IQR])	-	68 [56 - 77]	-	68 [56 - 77]	67 [55 - 76]	67 [53 - 79]
Sex (n (%))	Female	22,587 (50.96%)	-	99,673 (53.34%)	104,714 (48.46%)	22,570 (52.68%)
	Male	21,736 (49.04%)	-	87,096 (46.61%)	111,369 (51.54%)	20,271 (47.32%)
	Unknown	-	-	103 (0.06%)	-	-
Tinnitus						
Number records	-	145,877	64,915	456,222	319,831	189,384
Number subjects	-	119,169	45,881	317,074	152,834	185,622
Age (Median [IQR])	-	57 [45 - 67]	55 [42 - 65]	58 [48 - 69]	55 [43 - 65]	58 [46 - 70]
Sex (n (%))	Female	73,270 (50.23%)	31,266 (48.16%)	246,365 (54.00%)	151,225 (47.28%)	101,515 (53.60%)
	Male	72,607 (49.77%)	33,649 (51.84%)	209,649 (45.95%)	168,606 (52.72%)	87,869 (46.40%)
	Unknown	-	-	208 (0.05%)	-	-

12.2 Incidence rates

Outcomes classified as very rare if the maximum incidence rate was below 10 per 100,000 person-years, rare if between 10 and 100 per 100,000 person-years, or uncommon to common otherwise.

Figure 2. Standardised incidence rates per outcome stratified by database over full study period.

12.5.1 Immune-mediated diseases

Narcolepsy, Kawasaki disease, and Guillain-Barré syndrome were all very rare with incidence rates less than 10 per 100,000 person-years, see **Figure 2**. Type 1 diabetes, immune thrombocytopenia, and Bell's palsy had incidence rates less than 100 per 100,000 person-years, while autoimmune thyroiditis (narrow and broad) had higher incidence rates, although these were much higher in IQVIA DA Germany compared to other databases. The overall incidence rate for autoimmune thyroiditis (broad) was 122.01 (121.27 to 122.75) in IQVIA DA Germany, but ranged between 6.41 (6.22 to 6.61) and 43.62 (43.16 to 44.09) in the other databases.

Kawasaki disease, narcolepsy, and type 1 diabetes peaked in younger age groups. Autoimmune thyroiditis (narrow and broad) peaked in middle age, while Bell's palsy and Guillain-Barré syndrome were higher for older ages. Meanwhile, immune thrombocytopenia was higher for youngest and oldest ages, see **Figure 3**. Incidence rates for autoimmune thyroiditis (broad and narrow) were higher for females, while rates of Guillain-Barré syndrome were generally higher for males, see **Figure 4**. Time trends varied across databases. For example, Bell's palsy showed an upward trend in SIDIAP while Guillain-Barré syndrome showed a falling trend in IPCI, but such trends were not seen in the other databases, see **Figure 5**.

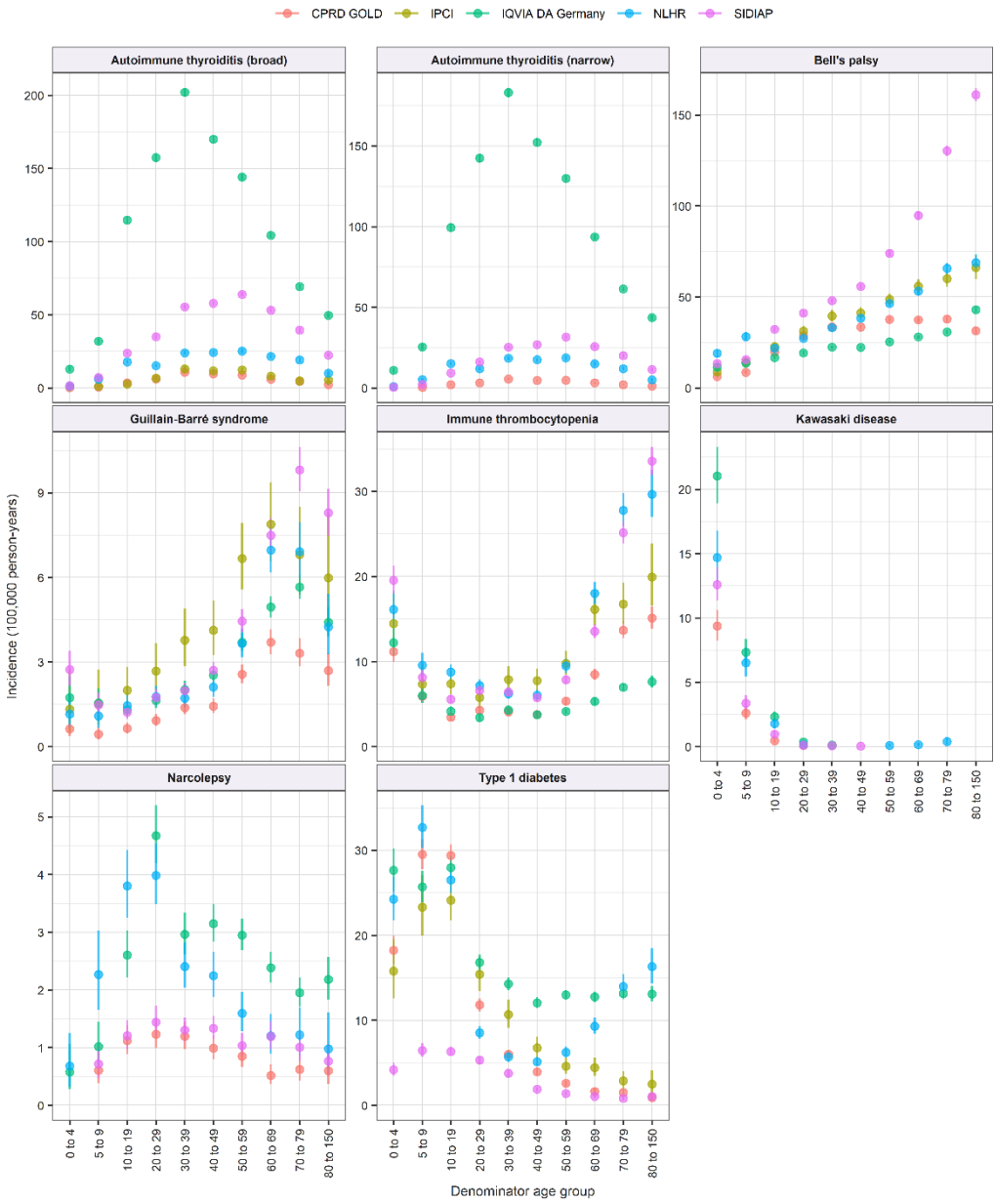

Figure 3. Incidence rates of immune-mediated diseases stratified by age.

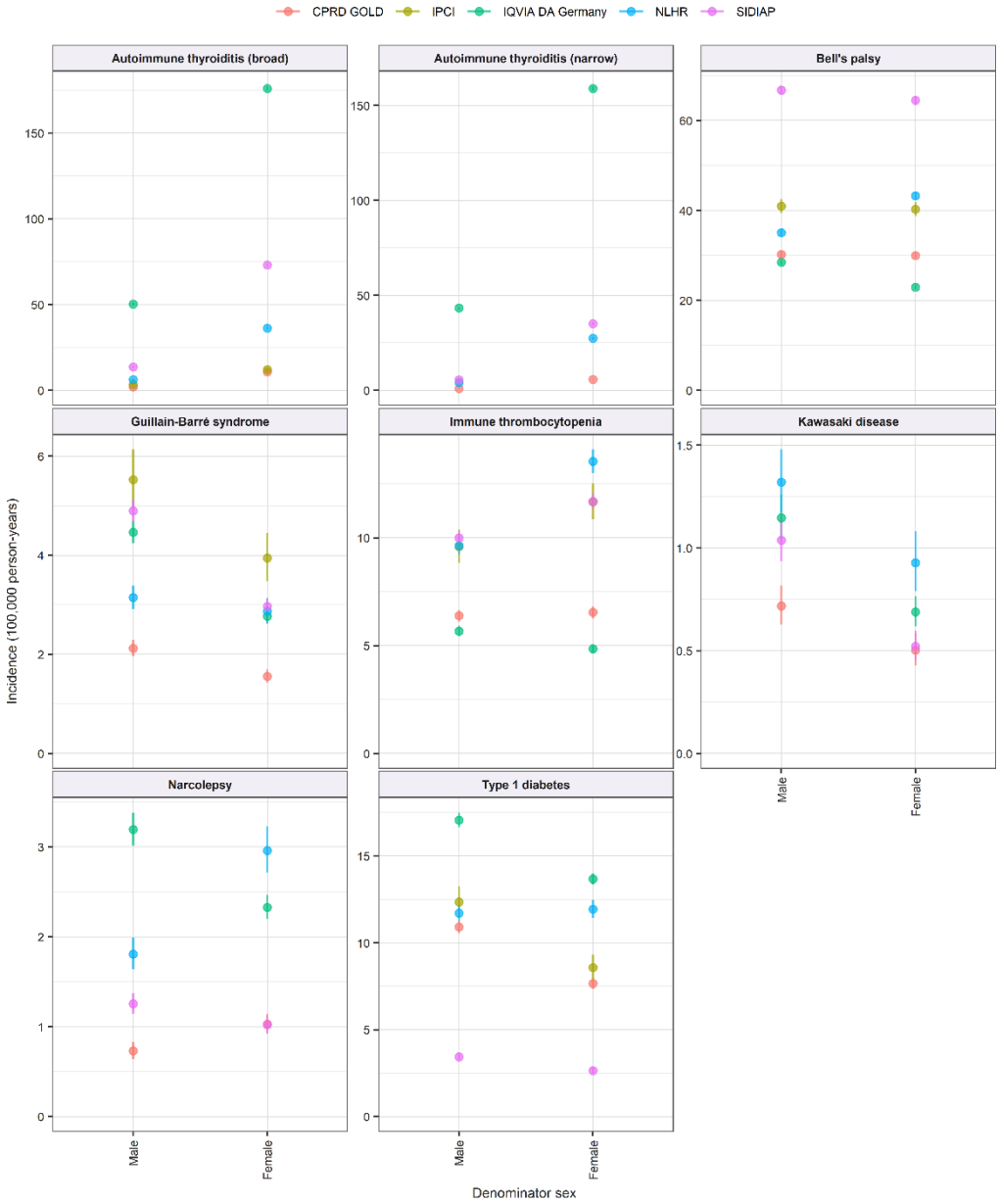

Figure 4. Incidence rates of immune-mediated diseases stratified by sex.

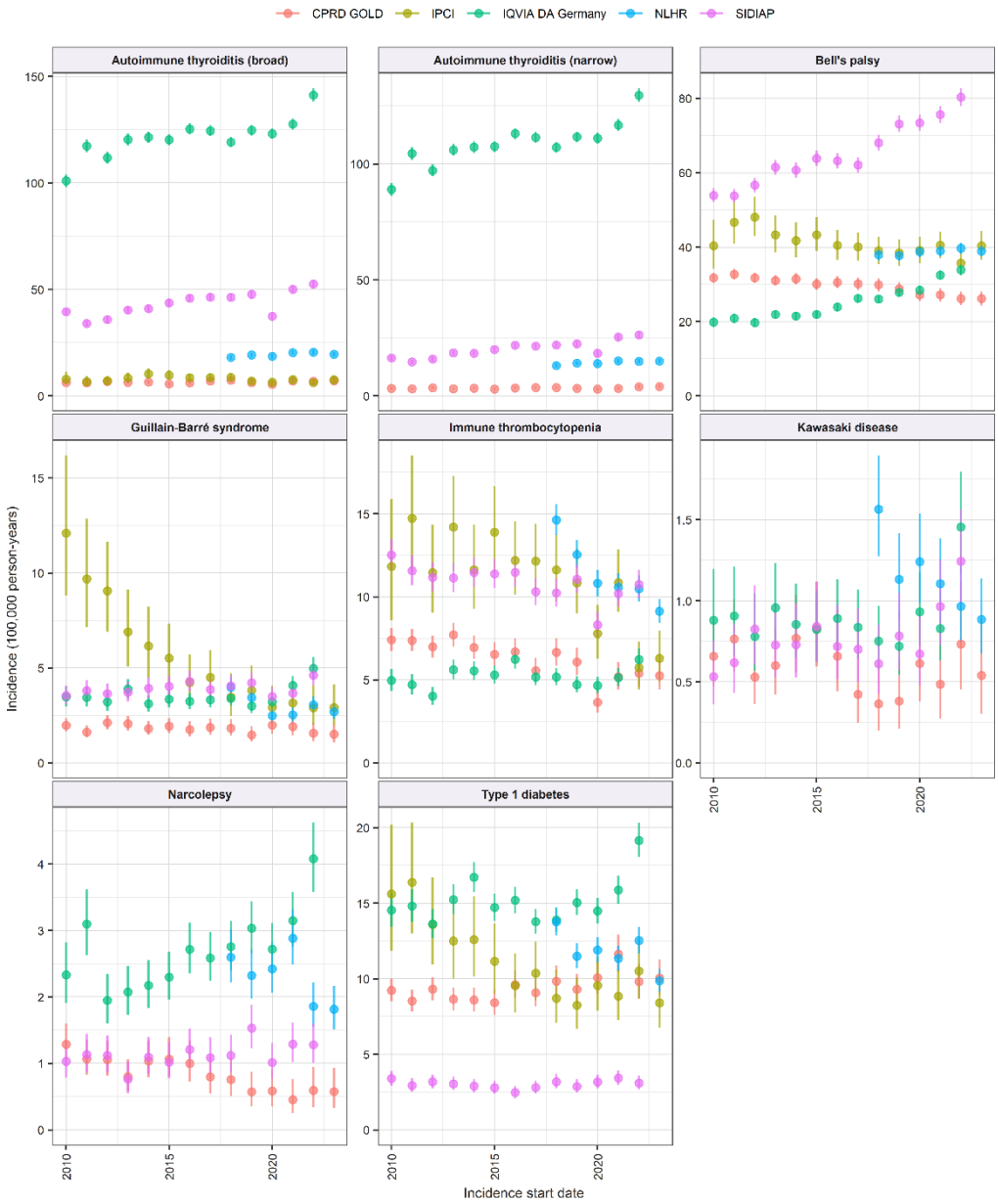

Figure 5. Incidence rates of immune-mediated diseases stratified by calendar year.

12.5.2 Blood disorders

Thrombotic thrombocytopenia syndrome (broad) and thrombotic thrombocytopenia syndrome (narrow) were both rare with incidence rates less than 10 per 100,000 person-years, **Figure 2**. Thrombocytopenia had higher incidence rates, but these varied substantially across databases. The lowest overall incidence rate for thrombocytopenia was 58.44 (57.70 - 59.20) in NLHR, while it was 284.76 (281.85 - 287.70) and 282.30 (281.18 - 283.42) in IPCI and IQVIA DA Germany, respectively. Meanwhile, estimates were 746.95 (744.85 - 749.06) in CPRD Gold and 1,327.01 (1,324.45 - 1,329.57) in SIDIAP.

Incidence rates for each of the cohorts increased with age, **Figure 6**, and were higher for males, **Figure 7**. Trends over time were heterogeneous across databases. Although there was a substantial increase in thrombotic thrombocytopenia syndrome (broad and narrow) seen in SIDIAP over time, such an increase was not seen in other databases, **Figure 8**.

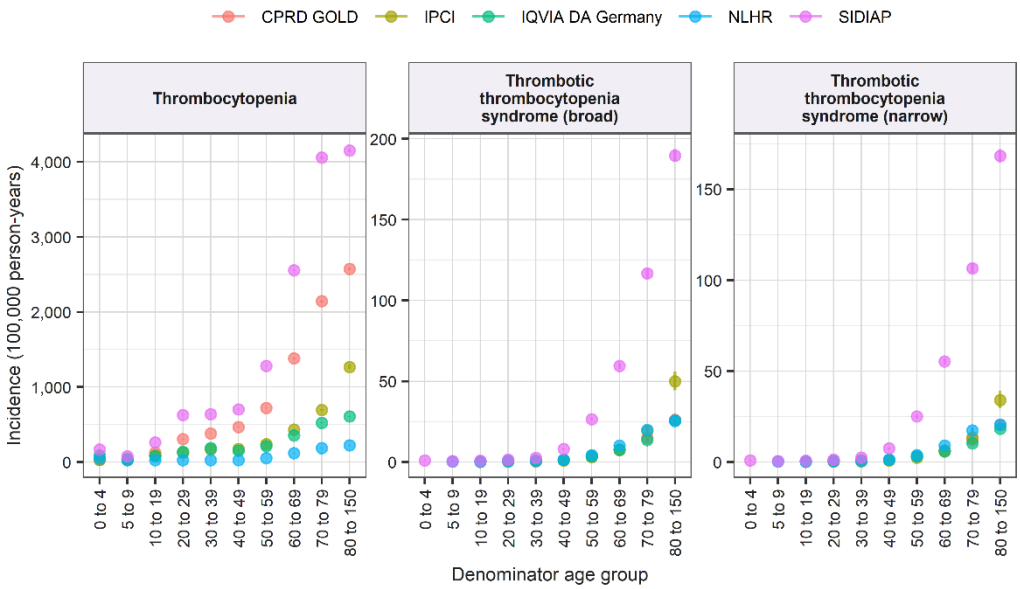

Figure 6. Incidence rates of blood disorders stratified by age.

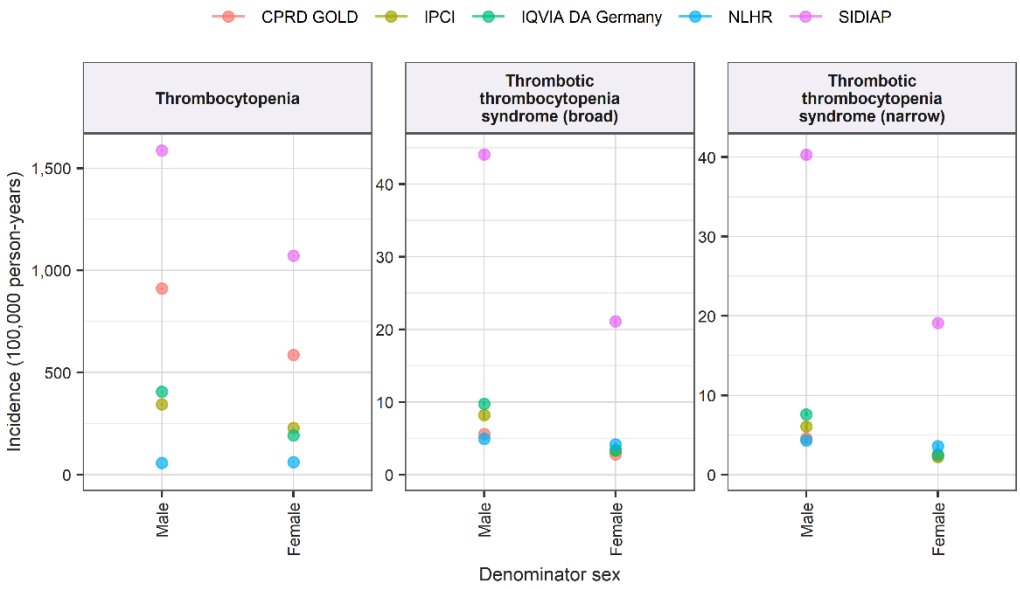

Figure 7. Incidence rates of blood disorders stratified by sex.

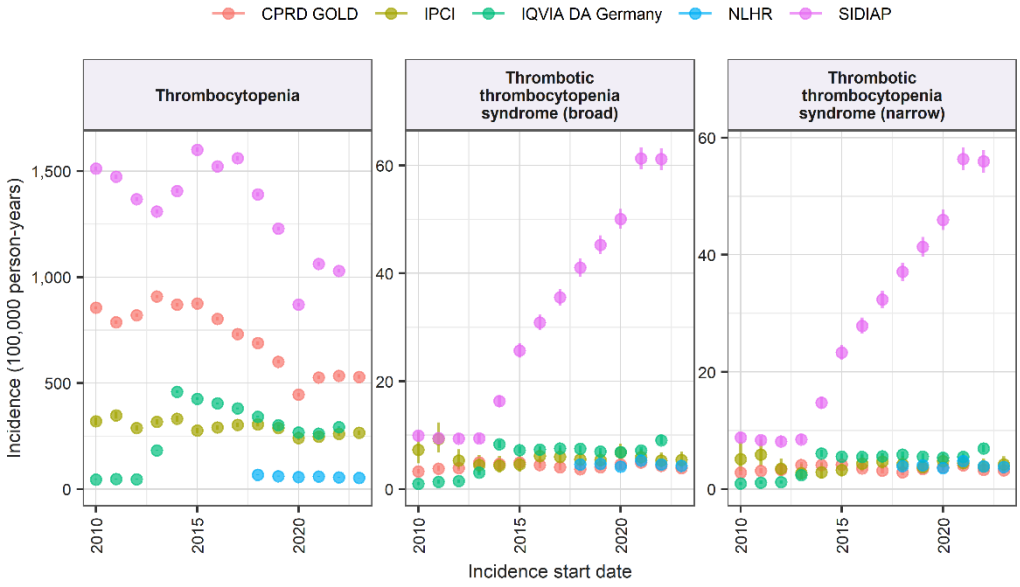

Figure 8. Incidence rates of blood disorders stratified by calendar year.

12.5.3 Disorders of the cardiovascular system

Thrombotic microangiopathy and single organ cutaneous vasculitis had the lowest incidence rates with incidence rates less than 10 per 100,000 person-years, see Figure 2. Cardiomyopathy (broad), cardiomyopathy (primary), IgA vasculitis, myocarditis, myocarditis or pericarditis, pericarditis (broad), and pericarditis (primary) had incidence rates less than 100 per 100,000 person-years. Arrythmia, coronary artery disease: myocardial infarction, coronary artery disease: all, coronary artery disease: angina, coronary artery disease: atherosclerosis, and heart failure were all relatively more common.

Incidence rates for almost all disorders of the cardiovascular system increased with age. The exceptions were IgA vasculitis and thrombotic microangiopathy which had peaks at both youngest and oldest ages. Meanwhile myocarditis peaked in middle age in IQVIA DA Germany, but was highest at oldest ages in NLHR, see **Figure 9**.

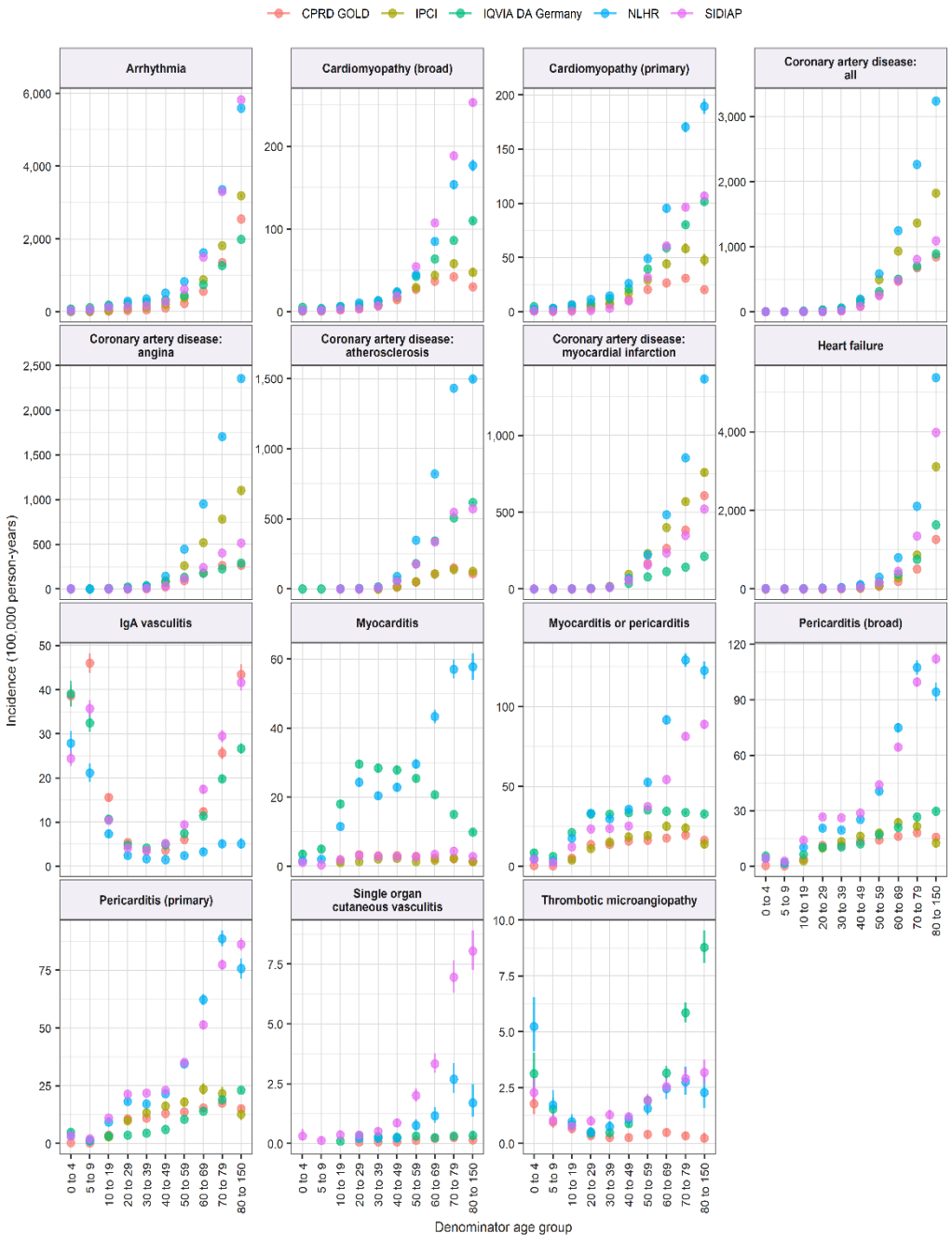

Figure 9. Incidence rates of disorders of the cardiovascular system stratified by age.

Figure 10. Incidence rates of disorders of the cardiovascular system stratified by sex.

Figure 11. Incidence rates of disorders of the cardiovascular system stratified by calendar year.

12.5.4 Disorders of the nervous system

Aseptic meningitis, encephalitis (primary), myelitis (narrow), myelitis (primary), and optic neuritis had the lowest incidence rates with incidence rates less than 10 per 100,000 person-years, see **Figure 2**. Encephalitis (broad), encephalitis (narrow), febrile seizure, and multiple sclerosis had incidence rates less than 100 per 100,000 person-years, while neuritis (broad), neuritis (narrow), neuritis (primary), non-epileptic seizure, and seizure were relatively more common.

Incidence rates for disorders of the nervous system were highest among youngest ages for aseptic meningitis, febrile seizure, and seizure. Encephalitis (narrow), encephalitis (broad), myelitis (narrow), and myelitis (primary) were highest for middle age in IQVIA DA Germany, but were higher at older ages in other databases. Optic neuritis peaked in middle-age, except for in SIDIAP where rates were highest for oldest ages. Multiple sclerosis was highest in middle-age in all databases, while incidence of neuritis was highest for older ages, see **Figure 12**. Incidence rates were mostly similar by sex, although rates for optic neuritis were higher for females, see **Figure 13**. Time trends were observed but varied by database, see **Figure 14**.

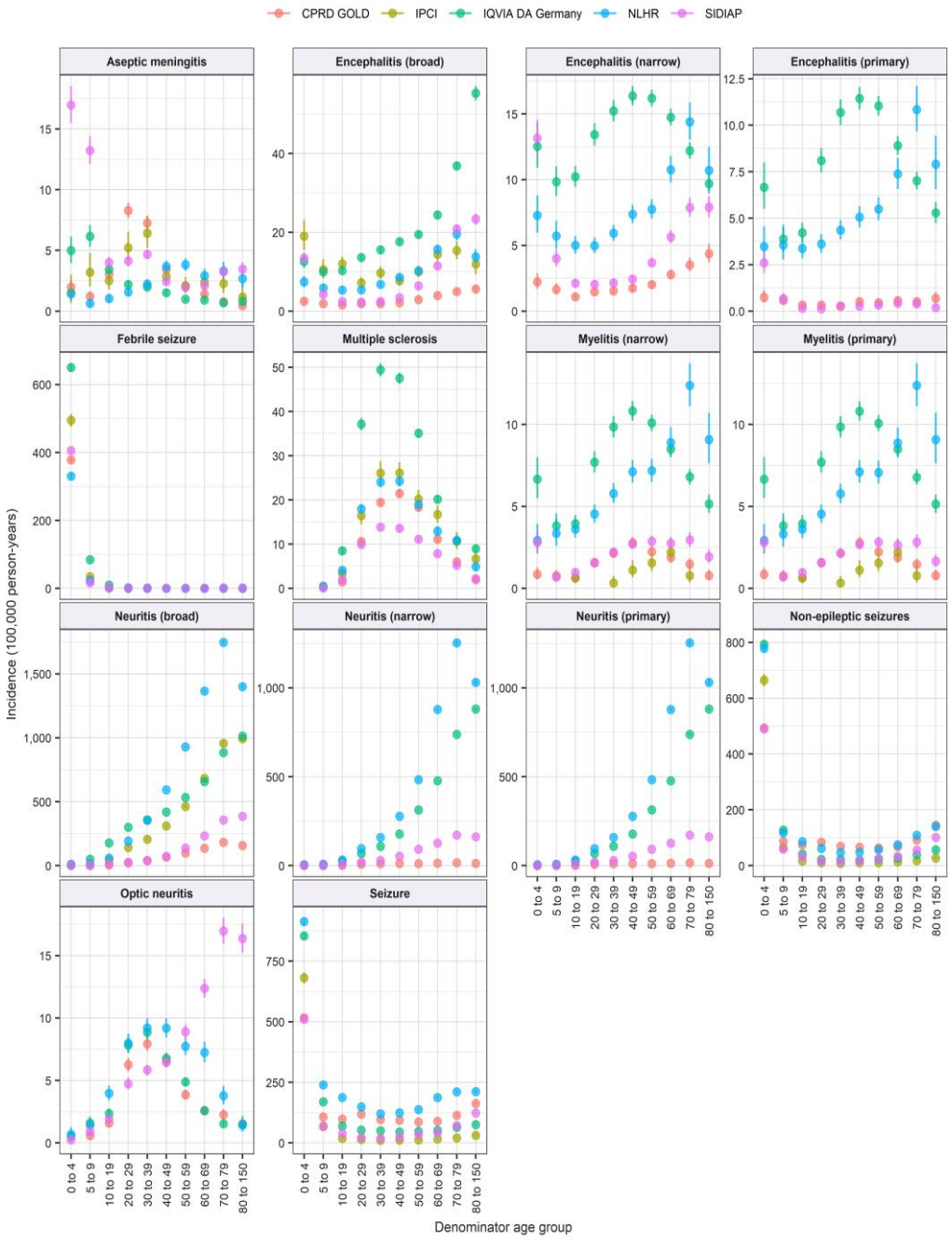

Figure 12. Incidence rates of disorders of the nervous system stratified by age.

Figure 13. Incidence rates of disorders of the nervous system stratified by sex.

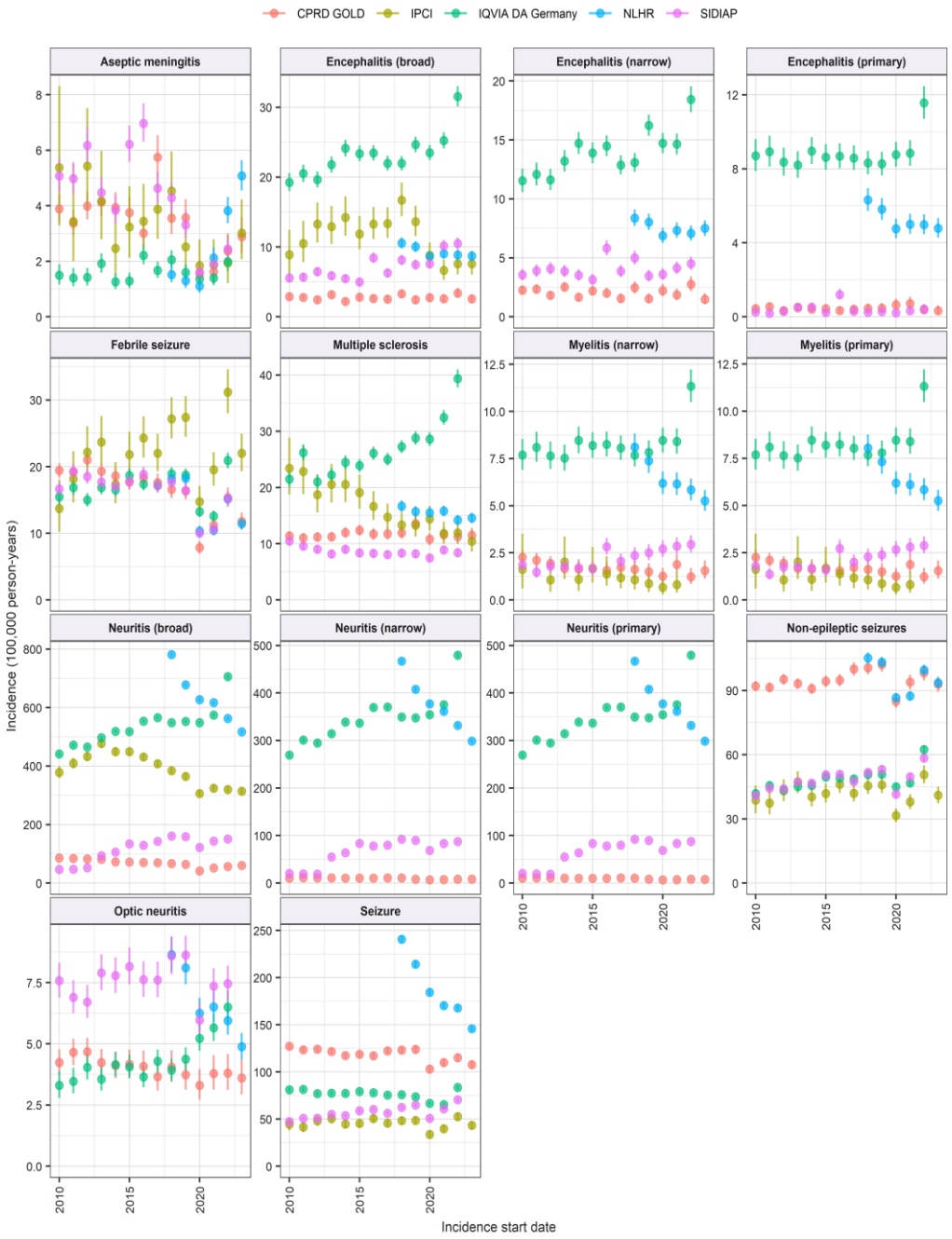

Figure 14. Incidence rates of disorders of the nervous system stratified by calendar year.

12.5.5 Coagulation disorders

Cerebral venous sinus thrombosis and disseminated intravascular coagulation had the lowest incidence rates with incidence rates less than 10 per 100,000 person-years, see Figure 2. All incidence rates for haemorrhagic stroke were less than 100 per 100,000 person-years, while deep vein thrombosis, ischaemic stroke, non-haemorrhagic stroke, pulmonary embolism (broad), and pulmonary embolism (primary) were all relatively more common.

Incidence rates for coagulation disorders increased with age, Figure 15. Rates of cerebral venous sinus thrombosis were generally higher for females, while rates of ischaemic stroke and non-haemorrhagic stroke cohort were higher for males, see Figure 16. Time trends were observed but varied by database, see **Figure 17**.

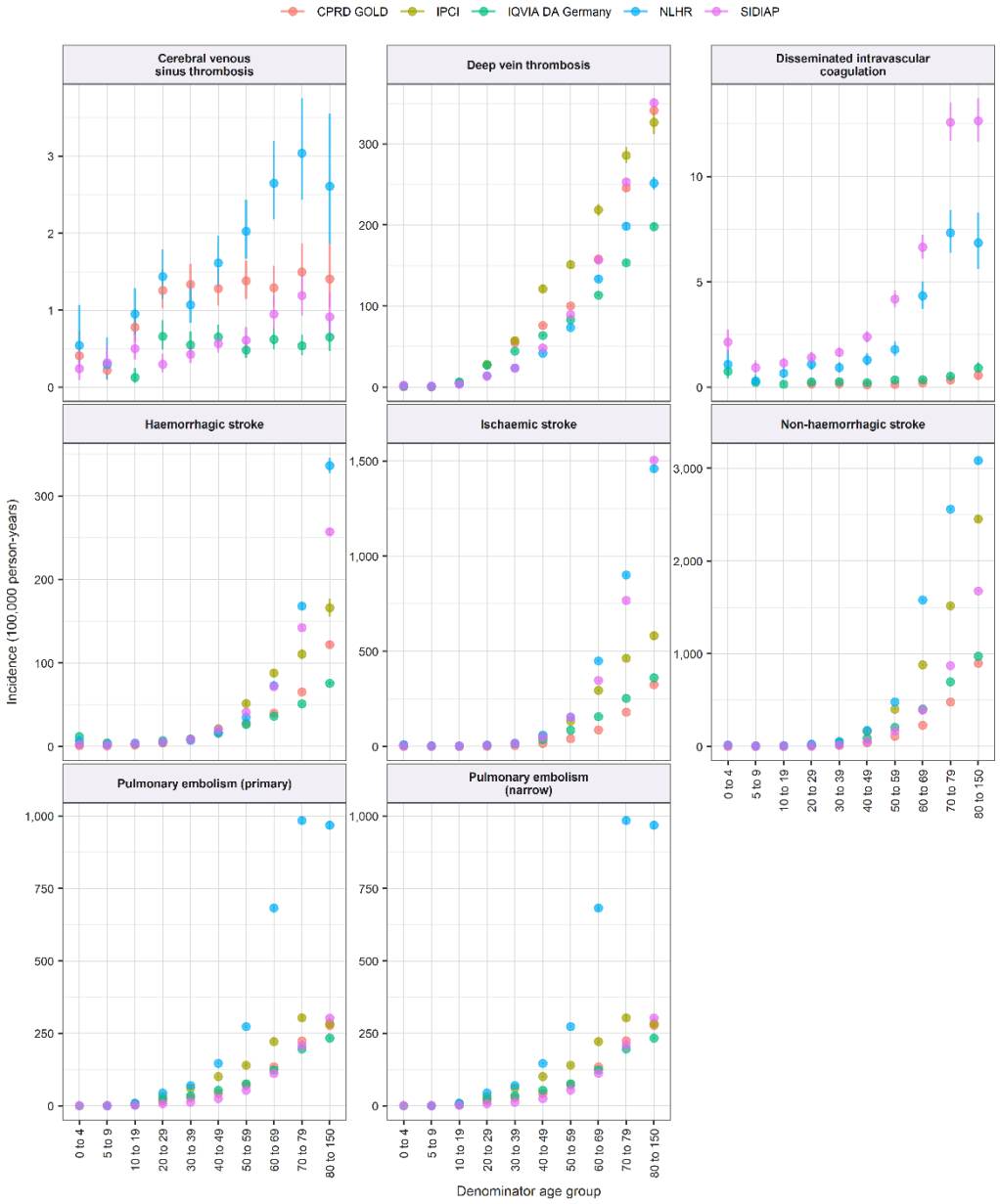

Figure 15. Incidence rates of coagulation disorders stratified by age.

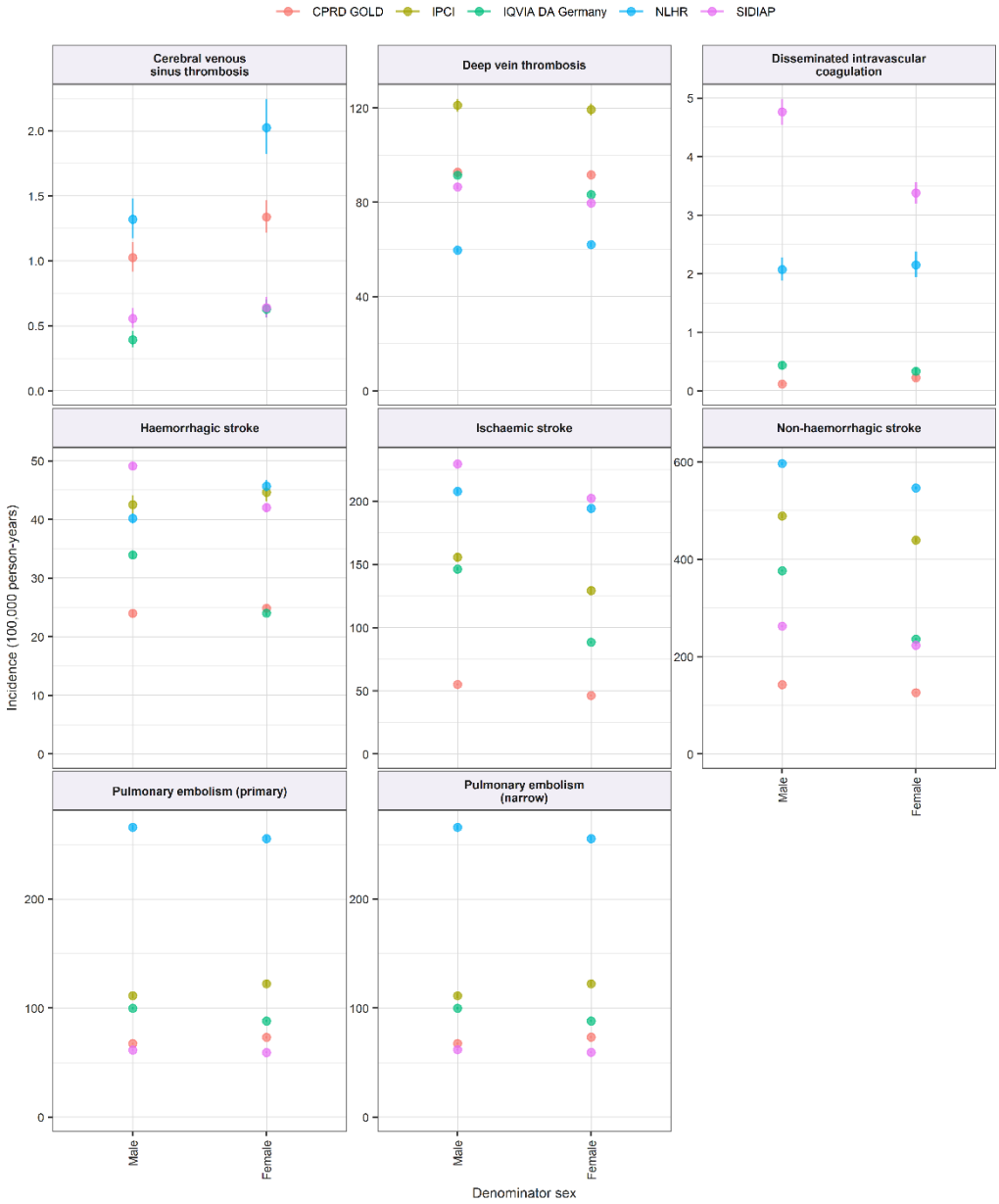

Figure 16. Incidence rates of coagulation disorders stratified by sex.

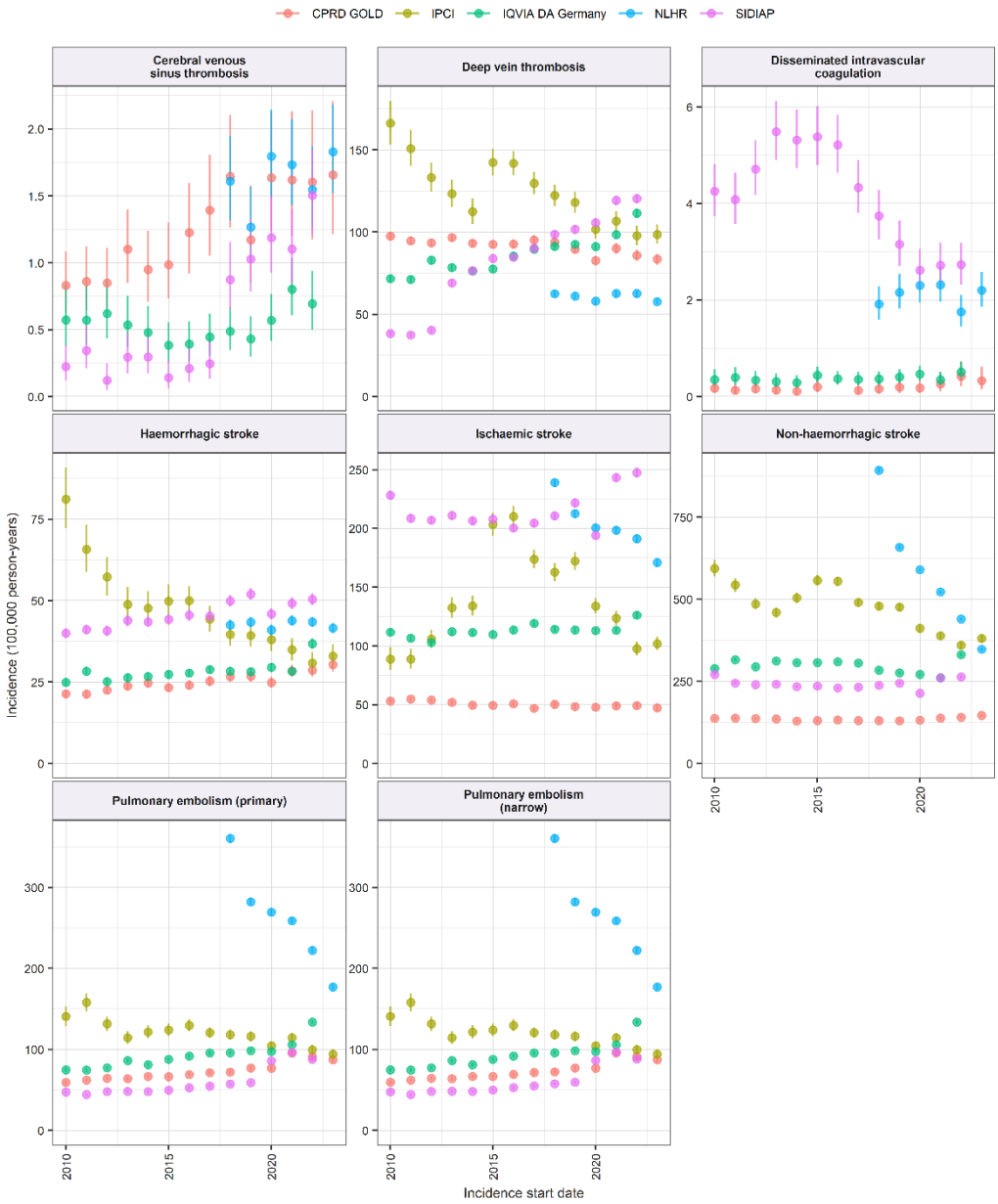

Figure 17. Incidence rates of coagulation disorders stratified by calendar year.

12.5.6 Disorders of the skin, bones and joints systems

Drug reaction with eosinophilia and systemic symptoms, and erythema multiforme and Stevens-Johnson syndrome/ toxic epidermal necrolysis all incidence rates less than 10 per 100,000 person-years, see Figure 2, while rheumatoid arthritis was relatively more common.

Incidence rates increased with age, except for erythema multiforme where rates were higher at youngest ages, **Figure 18**. Rates were also generally higher for females, **Figure 19**. Rates were generally stable over time, although incidence of erythema multiforme decreased over time in CPRD GOLD, **Figure 20**.

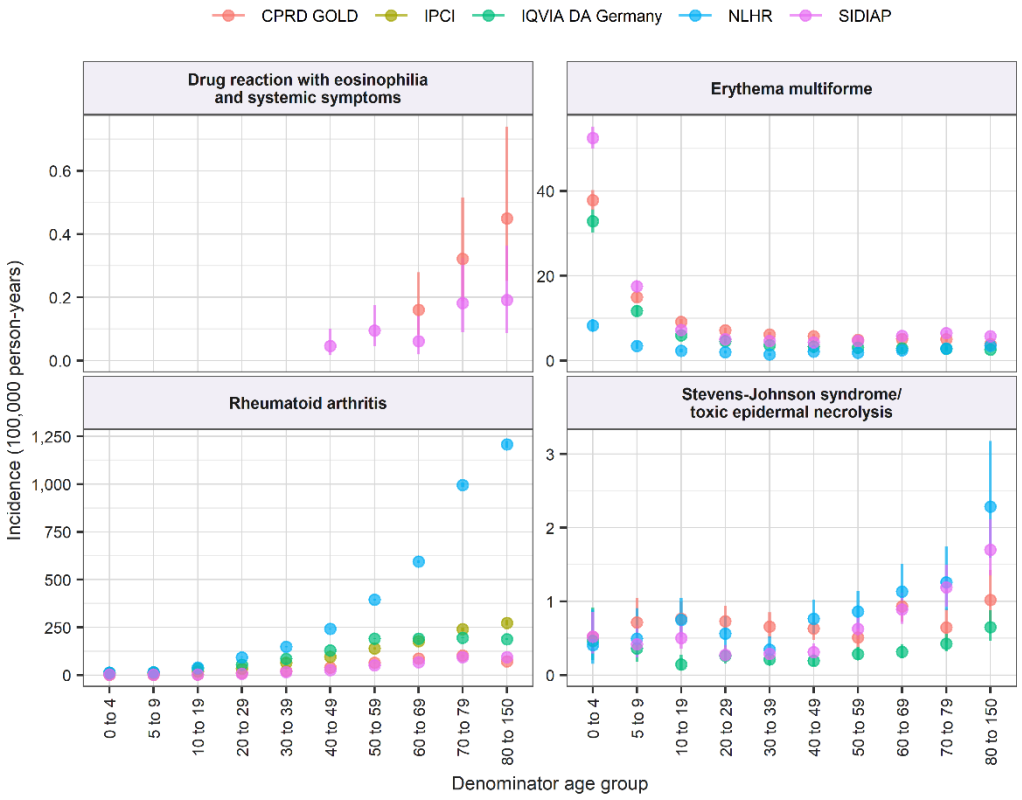

Figure 18. Incidence rates of disorders of the skin, bones and joints systems stratified by age.

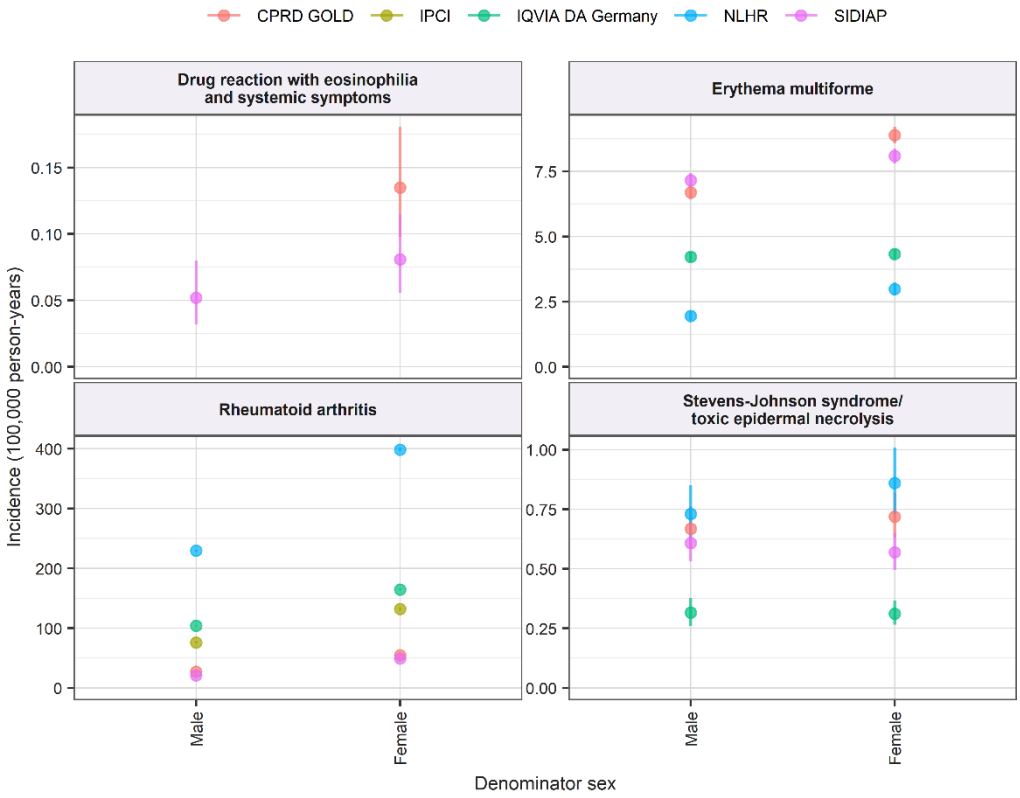

Figure 19. Incidence rates of disorders of the skin, bones and joints systems stratified by sex.

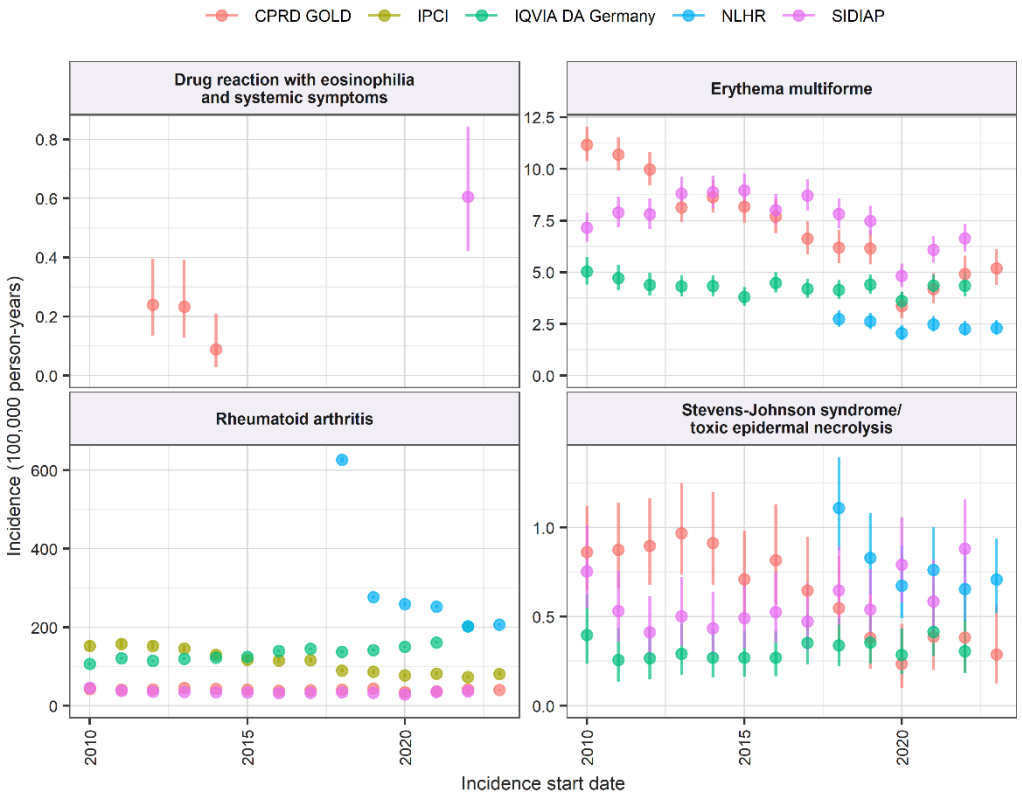

Figure 20. Incidence rates of disorders of the skin, bones and joints systems stratified by calendar year.

12.5.7 Disorders of other systems

Incidence rates for acute liver injury, acute pancreatitis (broad), acute pancreatitis (narrow), anaphylaxis, rhabdomyolysis (broad), and rhabdomyolysis (narrow) were less than 10 per 100,000 person-years, see Figure 2. Acute kidney injury (broad), acute kidney injury (primary), hearing loss, postmenopausal haemorrhage (broad), postmenopausal haemorrhage (narrow), sensorineural hearing loss, and tinnitus meanwhile were seen to be relatively more common.

Incidence rates for most disorders of other systems increased with age, although rates for postmenopausal haemorrhage were highest at the earliest qualifying ages, while rates for tinnitus peaked between age 50 and 70, see Figure 21. Rates for acute kidney injury, acute liver injury, and rhabdomyolysis were higher for males, Figure 22. Rates for acute kidney injury and rhabdomyolysis were seen to consistently increase over time, while other time trends varied by database, see **Figure 23**.

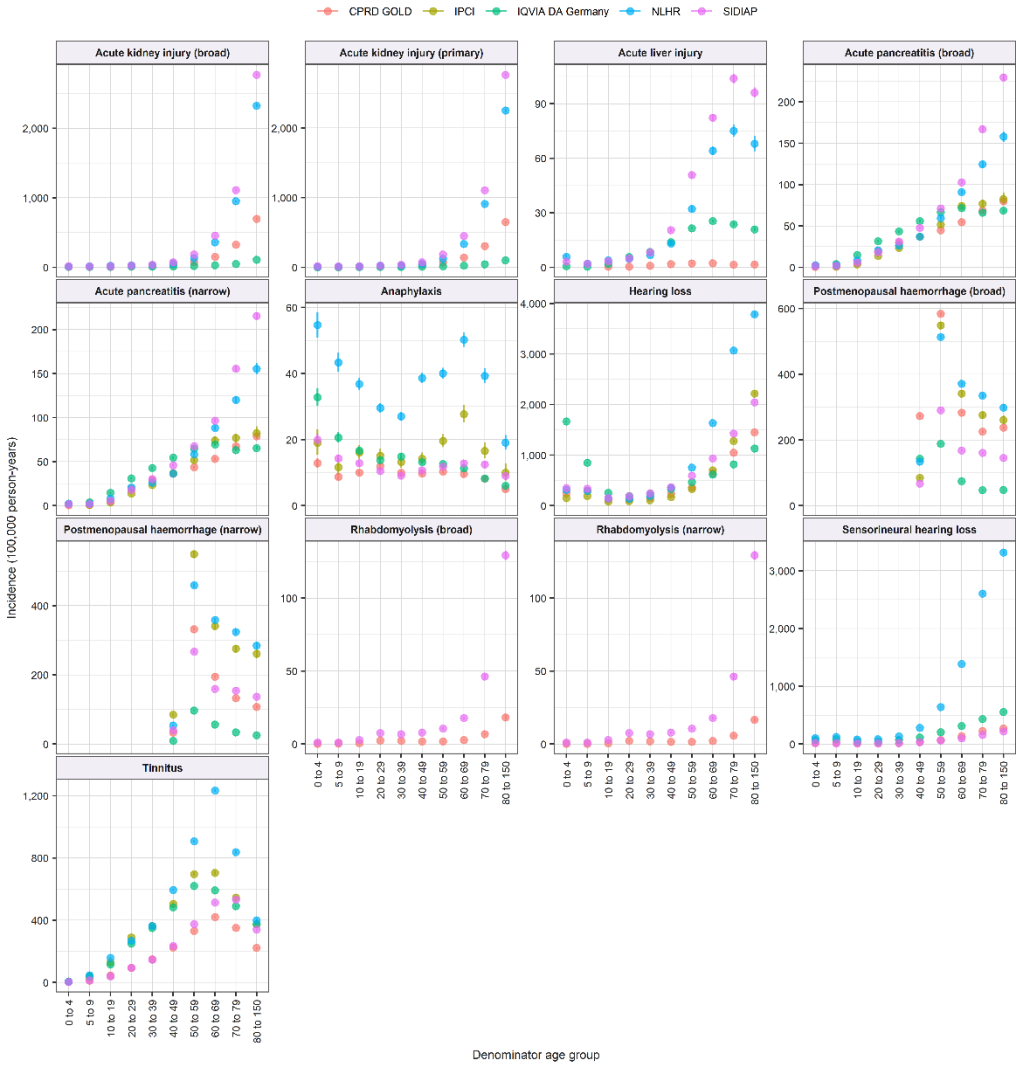

Figure 21. Incidence rates of disorders of other systems stratified by age.

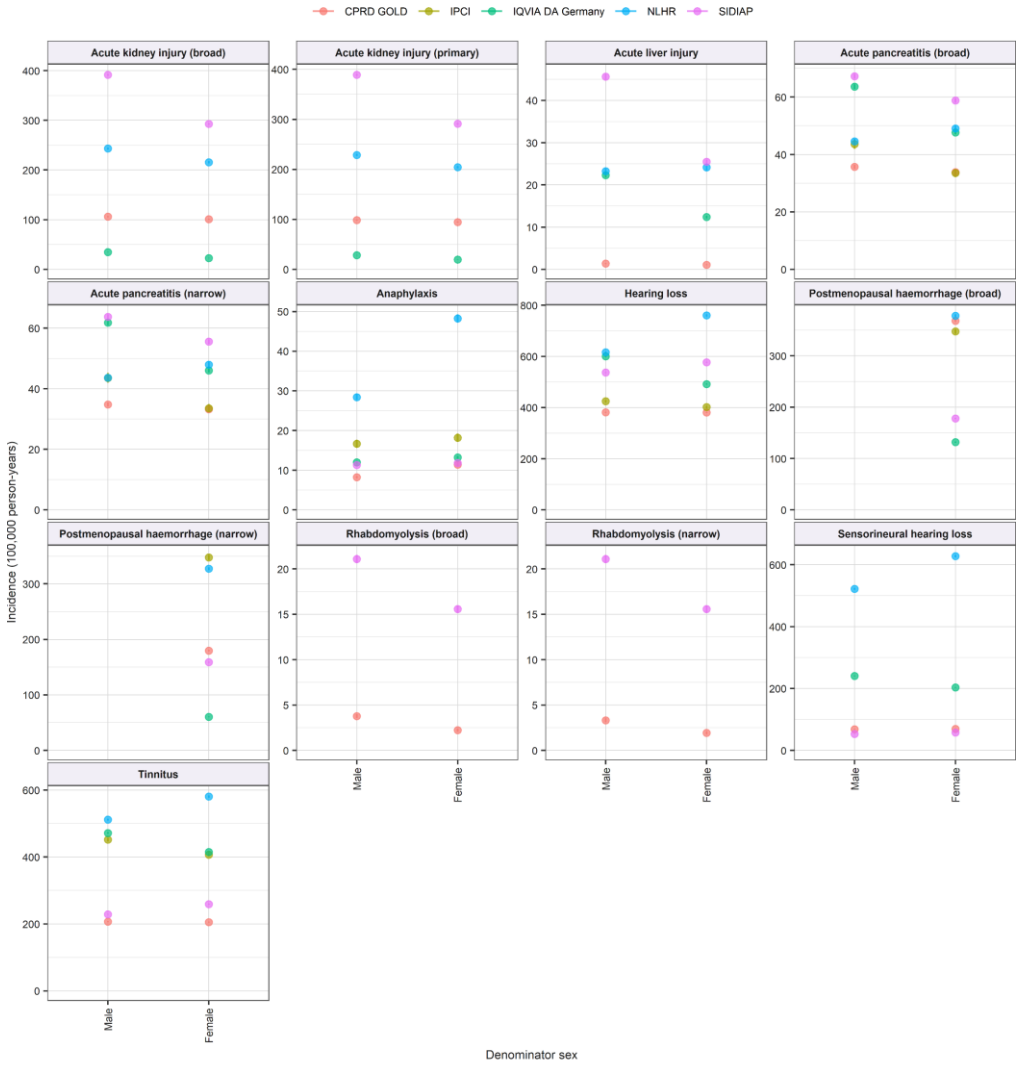

Figure 22. Incidence rates of disorders of other systems stratified by sex.

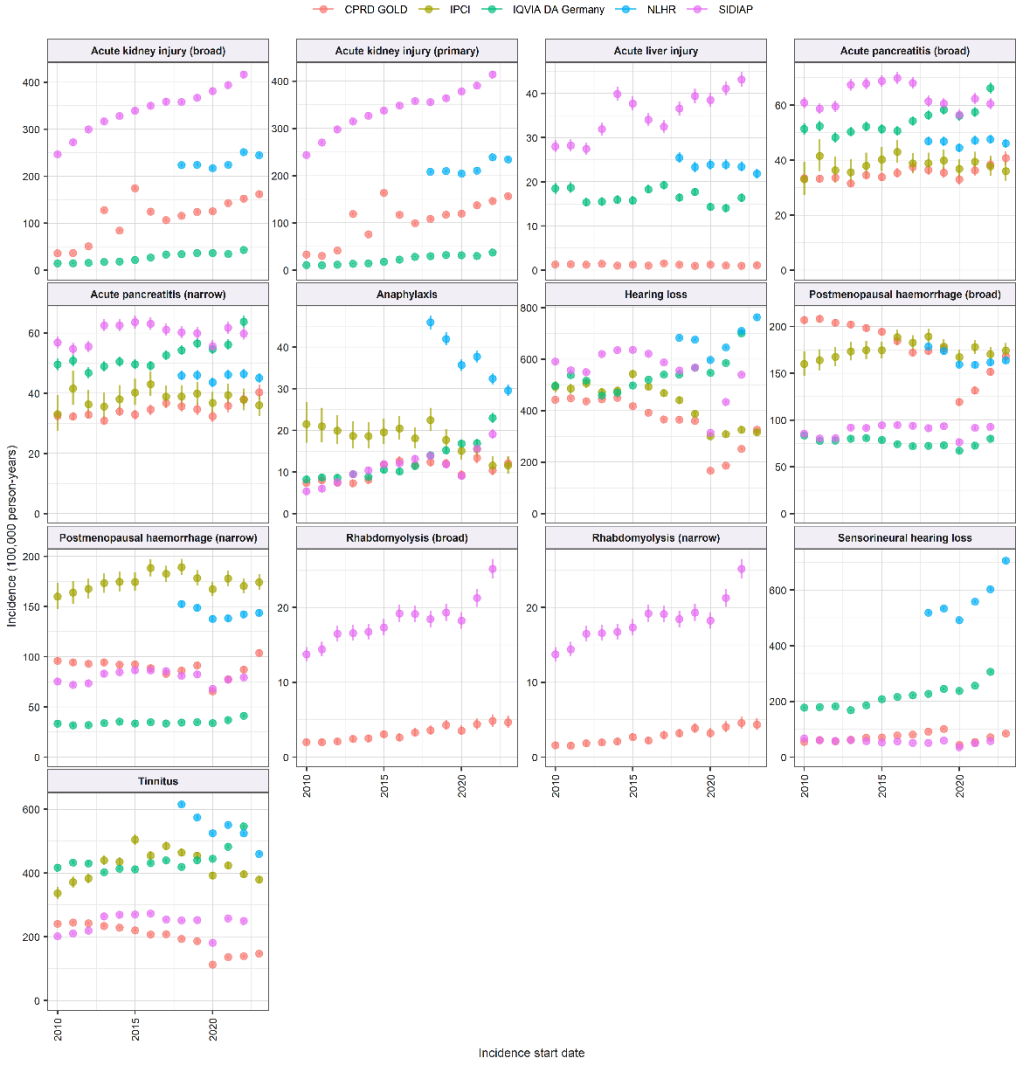

Figure 23. Incidence rates of disorders of other systems stratified by calendar year.

12.6 Characteristics of the AESI cohorts

Large scale characteristics for each of the AESI outcome cohorts are provided in the shiny web application, along with a comparison of the characteristics of individuals with a similar age and sex. As can be seen when comparing the characteristics of those individuals with an AESI to individuals with a similar age and sex, those with an outcome seem to generally have more prior comorbidities and medication use.

13. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Adverse events/adverse reactions were not collected in this study (not applicable as no product information is collected). The nature of this non-interventional study leveraging secondary use data does not fulfil the criteria for reporting adverse events, according to module VI, VI.C.1.2.1.2 of the Good Pharmacovigilance Practices (https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf).

Only in case of prospective data collection, there is a need to describe the procedures for the collection, management and reporting of individual cases of adverse events/adverse reactions.

14. DISCUSSION

14.1 Key results

Among immune-mediated diseases, autoimmune thyroiditis and Bell's palsy were the most common outcomes, while Kawasaki disease and Guillain-Barré syndrome were the rarest outcomes. Thrombocytopenia was the most common blood disorder, although estimates varied substantially across databases. Arrythmia, coronary heart disease, and heart failure were the most common disorders of the cardiovascular system, while thrombotic microangiopathy and single organ cutaneous vasculitis were the rarest. While neuritis was one of the most common disorders of the nervous system, optic neuritis was one of the rarest. Among coagulation disorders, non-haemorrhagic stroke was most common, while cerebral venous sinus thrombosis and disseminated intravascular coagulation were the rarest outcomes. Meanwhile for disorders of other systems, tinnitus and hearing loss consistently had the highest incidence rates.

Most outcomes had higher incidence for older age groups, although Kawasaki, type 1 diabetes, and seizures were more common in the youngest age groups, autoimmune thyroiditis, multiple sclerosis, optic neuritis, and tinnitus peaked in middle age, while immune thrombocytopenia, thrombotic microangiopathy, and IgA vasculitis were more common in the youngest and oldest age groups. In general differences were less pronounced when stratifying by sex, although outcomes such as autoimmune thyroiditis and optic neuritis were common among females while outcomes such as Guillain-Barré syndrome, acute kidney injury, acute liver injury, and rhabdomyolysis were more frequently seen for males. Many outcomes had notable time trends in incidence rates, however these were mostly inconsistent and differed by database. Where different definitions for the same outcome were considered, the most pronounced difference in estimates was seen for neuritis, cardiomyopathy, and encephalitis.

When comparing the characteristics of those individuals with an AESI to individuals with a similar age and sex, those with an outcome were observed to have more comorbidities and prior medication use.

For some of the study outcomes especially chronic conditions or with extended progression over time, there was evidence of index date misclassification. This should be taken into consideration for future studies where this is particularly important, such as when using self-controlled methods.

14.2 Limitations of the research methods

General limitations:

The results estimated from this study will only reflect the populations from the included data sources. Electronic health records have certain inherent limitations because they were collected for clinical purpose rather than primarily for research use. We assume that if there were no related clinical codes of a condition presented for an individual in the data, the condition does not exist for this individual.

Misclassification of outcomes could happen if individuals received health care service outside of the data capture system. For example, in the UK and IPCI primary care data, we would not be able to capture event recorded in private care sectors. All the selected data sources are representative of the general source population and are stable over time. Therefore, the potential impact of misclassification is expected to be similar across the data sources throughout the study period. The data source setting may also impact capture of diagnosis, therefore data sources that record only primary care diagnosis such as CPRD GOLD may underreport diagnoses made in hospital. For example, diagnosis and tests of serious and emergency conditions like GBS is typically conducted in hospital setting.

Study-specific limitations:

While we developed the phenotypes for all study outcomes using the standard procedure, as well as conduct the diagnostics in the participating databases, these phenotypes may not fully apply to other databases, and further diagnostics would be needed when applying these phenotypes in other databases or later versions of the same databases.

Since published literature was not available for all AESIs to determine the appropriate length of the clean period, it is possible that some periods were set incorrectly and may lead some rates to reflect a combination of prevalent and incident cases, especially for outcomes with extended disease progression. In future studies, it is important to first check the patient characteristics in each database before implementing a phenotype to analysis. Methods to mitigate the impact such as conducting sensitivity analysis of different clean periods, using proxies of conditions including treatment use, and restricting to certain patient group based on clinical knowledge can be considered.

Additionally, changes in clinical guidelines or practice of recording of the electronic health records could affect the estimation of incidence rates over time.

14.3 Interpretation

Comparing to existing literature:

Since the start of the COVID-19 pandemic, several studies have been conducted and published on estimating the background rates of potential AESIs for COVID-19 vaccines using different data sources from multiple countries across the world. These include multi-databases studies [7,8,10,20] and studies conducted within individual countries or regions such as the US, Canada, Sweden, Australia, Scotland, and Hong Kong. [21–28]

For example, the EMA-funded ACCESS (The vACCine covid-19 monitoring readinESS) project generated background incidence rates of 41 AESIs with 10 databases from 7 European countries (Italy, Spain, Denmark, The Netherlands, Germany, France and United Kingdom) with a distributed network approach. [8] Within the OHDSI network, Li et al. estimated the background incidence rates of 15 prespecified AESIs associated with covid-19 vaccines during 2017 to 2019. The study included 13 databases from eight countries (Australia, France, Germany, Japan, the Netherlands, Spain, the United Kingdom, and the United

States) and all databases were mapped to the OMOP CDM. [7] Another multinational study was led by the Global Vaccine Data Network (GVDN) consortium with 11 databases from nine countries and regions (Denmark, England, Finland, France, Scotland, Taiwan, Argentina, Australia, Canada). A total of 13 pre-defined AESIs were included with the study period of 2015 to 2020.

Compared to previous studies, the current study covered a longer study period (2010 to 2023), which included both pre- and post-pandemic periods, allowing to also see the underlying time trend of the incidence rates. The current study included a broad list of AESIs, which was not limited to AESIs associated with COVID-19 vaccines. The most recent background rates estimated from this study would be useful for future safety surveillance. Apart from the background rates, the current study also provided detailed characteristics of the cohorts, and contextualised the cohort characteristics with the matched general population using varied time windows prior to index. To increase transparency and reproducibility of phenotypes use, we reported the use and counts of all clinical codes, in both OMOP standard concept and in source data, in each database to help interpret the results.

Heterogeneous of results from multi-database studies have been observed from the current study as well as other previous studies, including the ACCESS project, OHDSI study, and the GVDN project. Heterogeneity could result from true differences in database characteristics, including the region and size of the study population, type of data source/health care system and availability of linkage (primary care and/or hospital data), the use of different coding systems, differences in health care and coding practices (e.g. universal healthcare with GP vs. private care, clinical care vs reimbursement purposes), type of data sources (e.g. electronic health records vs. registry data such as NLHR).

Using the age and gender stratified background incidence rates from for the years 2017 to 2019 estimated in two studies, Russek et al. explored how the different between-database sources of heterogeneity influenced the background rates estimation of venous thromboembolism. [8,20,29] The study stated that substantial heterogeneity in the background IRs was observed between all included data sources, in addition to observed within-data-source differences across age groups and genders. Using forest plots and random-effects models, the study found that databases collecting data from different parts of the health-care systems were the largest contributors to heterogeneity in estimates.

While direct comparison is difficult due to the differences between data sources, study period and age groups stratification, the background rates of many conditions estimated from the current study were comparable with estimates from the literature. Most of the databases included in the current study had been used in other reports, which allowed us to compare the incidence rates within the same database. For example, CPRD GOLD and SIDIAP data were used in both the OHDSI study and the ACCESS project, and IPCI was included in the OHDSI study. The NLHR data were included in the study by Pottegard et al. in an observed-to-expected analysis.[30]

For immune-mediated diseases, incidence rates of narcolepsy, GBS, bells' palsy, and immune thrombocytopenia were similar to the published literature.

In our study, for Type 1 diabetes, the incidence rates among the 0-19 years old group were consistent across all databases except in SIDIAP (6 per 100,000 person-years in SIDIAP and 22 to 28 per 100,000 person-years in other databases). The estimated incidence rates then decreased or plateaued for older age groups. In the ACCESS project, incidence rates were higher than the current study, and showed an increase trend with age in some of the database. For example, in CPRD, the reported IR among the 0-19 and 20-29 were 38 and 41 per 100,000 person-years, respectively. Despite the difference in estimates, our results were more in line with other research. A recent systematic review and meta-analysis pooled incidence rates from over 100 studies, and reported that the IR for type 1 diabetes among western Europe was 22.45 (19.55 to 25.79) per 100,000 person-years among children and adolescents under 20 years old.[31]

Incidence rates of thrombocytopenia were much higher in the current study than estimates from ACCESS and the Scandinavian study by Pottegard et al. For example, the overall incidence rates ranged from 18 to 92 per 100,000 person-years for different provenance of the events databases in ACCESS, 15 to 38 per 100,000 person-years in the Scandinavia study. In the current study, the overall incidence rates ranged from 28 (NLHR) to 1327 (SIDIAP) per 100,000 person years. The difference was driven by the inclusion of platelet count in the definition used by the current study, whereas only diagnosis codes were used in ACCESS. Consequently, data sources with hospital linkage and laboratory results such as SIDIAP had higher incidence rates. Finally, differences in socio-demographics could also explain some of the observed differences, with for example the Scandinavian study was limited to people aged 18 to 65.

Background rates of coagulation disorders including venous thromboembolism have been estimated in many previous studies, and rates of VTE varies across literatures. For example, in ACCESS study, the incidence rates of VTE were 141 to 229 per 100,000 person-years for the overall population, and from 510 to over 1,000 per 100,000 person-years among those aged over 80, depending on type of databases. In the OHDSI study, rates were only available with age-sex stratification. For those aged over 85, incidence rates ranged from 380 to 1100 per 100,000 person-years for deep vein thrombosis, and 105 to 365 per 100,000 person-years for PE among the European databases. In the GVDN study, incidence rates of pulmonary embolism ranged from 45 to 95 per 100,000 person-years for the overall population.

Background rates of deep vein thrombosis from the current study were consistent with some of the literature. For example, Pottegard et al. reported that the incidence rates of deep vein thrombosis was 48 per 100,000 person-years in the Norway data, and the corresponding rate in our study was 45 per 100,000 person-years in NLHR. The incidence rates of cerebral venous thrombosis and disseminated intravascular coagulation were also similar to the literature. Our estimates of background rates of pulmonary embolism were similar to the estimates in the OHDSI study and the Scandinavia study except in NLHR. In the current study, incidence rates of pulmonary embolism were much higher in NLHR compared to other included databases, especially among the older age groups.

Some of the study outcomes were not included in other research estimating background rates, such as post-menopause bleeding, rhabdomyolysis, and tinnitus. A prospective cohort study from Denmark reported an incidence of 409 and 42 per 1,000 person years during the first 12 months and more than 3 years after menopause, respectively. [32] Our estimates were much lower compared to that study, suggesting that postmenopausal bleeding may not be well recorded in routinely collected data or that the phenotype should be improved. We only identified patients with rhabdomyolysis in two databases (CPRD GOLD, SIDIAP), and we did not find much literature on the incidence rates of rhabdomyolysis among the general population. A systematic review and meta-analysis showed that the pooled incidence rate of any tinnitus was 1164 per 100,000 person-years (95% CI: 479-2828) using 12 studies globally.[33] A UK-based study used CPRD data and reported age-standardised incidence rate of 250 new tinnitus cases per 100,000 person-years (95% CI: 246–255). [34] The incidence rates of tinnitus estimated from the current study were consistent with that study.

The age and/or sex patterns of background rates observed in this study were generally consistent with other studies. For example, we observed that incidence rates of cardiovascular events and coagulation disorders increased with age. We also observed different time trends for different events, suggesting that when background rates are used in safety signal detection, it is important to use estimates from a more recent time period and to consider calendar time comparisons.

Interpretation of phenotype:

While we conducted standardised and reproducible phenotyping using distributed analytics and tools following the Darwin-EU procedure, phenotypes from this study had some limitations and needed to be

interpreted and used with caution. The limitations of phenotypes can be summarised into four categories: a) low sensitivity, which means we failed to identify some patients; b) low specificity, in which case we cannot differentiate some conditions to others mainly due to granularity of vocabularies, c) misclassification of index date, that we may not capture the true incident time, mainly due to delayed or incomplete diagnosis records; and d) other limitations such as unspecific clinical descriptions that were broad and included conditions very different aetiology.

For acute kidney injury, the phenotype is limited by low sensitivity. First, we did not use measurements to define AKI, whiles definition of AKI by clinicians is typically based on increase in proteinuria or serum creatinine vs previous/baseline. This will lead to underestimation and likely bias to more severe cases. We did not see any cases in IPCI, which likely because their most common code does not differentiate chronic vs acute renal disease. Besides, the linkage to hospital likely makes a difference, for example in SIDIAP the rates are ~3x fold CPRD GOLD's.

For acute liver injury, the phenotype was of low specificity as we cannot differentiate well acute from chronic liver injury other than in CPRD GOLD. The code of Acute hepatic failure is only found in CPRD GOLD data.

For many outcomes, misclassification of index date is likely, and this was observed when checking the characteristics of the cohorts.

For anaphylaxis, we observed some epinephrine prescriptions 1d to 1 year before index date, which suggested that misclassification of index date could exist.

For arrhythmia, we used condition and observation codes to identify the cohort, without including procedure codes related to arrhythmia. While this could reduce the sensitivity, including treatment procedure may lead to index date misclassification.

For heart failure, index date misclassification could exist as we observed about 10% of patients already on treatment with e.g. diuretics in the month and year before diagnosis.

For multiple sclerosis, misclassification of index date exists as the age onset in the current study was higher than literature. For pericarditis, characteristics showed that some patients experienced symptoms in the previous month (chest pain, dyspnea), ECG, auscultation, and some treatments (e.g. colchicine almost 9%).

For febrile seizure, the phenotype could be low specificity as we did not exclude those with a concomitant central nervous system infection, i.e. children with concomitant diagnosis or meningitis, encephalitis, meningoencephalitis. Measurement of temperature >38C was not included in the definition.

Phenotype of rheumatoid arthritis also has index date misclassification, with 10-15% patients on DMARDs treatment in the year before index.

In the single organ cutaneous vasculitis cohort, some individuals with systemic lupus erythematosus with organ/system involvement at index date, suggesting that the cutaneous vasculitis could be associated with underlying diseases. Additionally, on the index date, some individuals were already on corticosteroids, such as prednisolone, prednisone, and hydrocortisone. This suggests potential misclassification of the index date.

For type 1 diabetes, though we implemented multiple exclusion criteria, misclassification of index date still exists. For example, we observe higher age at index in the IQVIA DA Germany data as compared to others. This could because the IQVIA DA Germany database included general as well as diabetologist practices. Therefore, some patients were only captured when they visited these practices, and the diagnosis codes reflected a historically diagnosed condition. In a study looking at outpatients with type 1 diabetes using the

same data source, the mean and standard deviation of age was 45.3 (16.7) among the included patients.[35]

14.4 Generalisability

This study was run against a variety of sources of routinely-collected health care data from across Europe all mapped to a common data model. When considering relative trends in incidence rates by age and sex, consistent patterns were seen across the different data sources. These relative trends can therefore be expected to be generalisable outside of these data sources and indeed are consistent with the literature. However, absolute estimates often varied by data source. Moreover, for a number of outcomes there were differing trends over time. Consequently, the generalisability of database-specific estimates can be expected to be somewhat limited, and where possible comparisons between expected rates and observed rates should be made within the same data sources and also account for underlying time trends in that data source.

The phenotypes developed for this study were created specifically for those data sources included, and for the purpose of estimating background rates of AESIs of vaccines. To re-use these in other data sources and/or other studies, it is important to first have clear clinical descriptions of the outcomes of interest. It should be preceded by running diagnostics to make sure that no changes in definitions would be needed.

14.5 Other information

Not applicable.

15. CONCLUSION

This study included a wide range of adverse events of special interest for vaccines. We estimated background rates by year, age, and sex for five European databases. We also provided detailed cohort characteristics among people with the conditions, and contextualised the results by comparing to the matched cohort from the general population. However, the background rates need to be interpreted with caution given heterogeneity across databases and underlying time trends seen for many of the outcomes.

For any new studies aiming at using background rates for an emerging signal evaluation, it will be important to first assess if the phenotypes are fully aligned with the outcome(s) to be assessed, run diagnostics in the databases, and tailor as needed (e.g., considering information from spontaneous case reports and clinical case definitions). This work establishes a framework for future studies.

16. REFERENCES

1 Black SB, Law B, Chen RT, *et al.* The critical role of background rates of possible adverse events in the assessment of COVID-19 vaccine safety. *Vaccine*. 2021;39:2712–8.

2 van der Boom MDX, van Eekeren R, van Hunsel FPAM. Observed-over-Expected analysis as additional method for pharmacovigilance signal detection in large-scaled spontaneous adverse event reporting. *Pharmacoepidemiol Drug Saf*. 2023;32:783–94.

3 Mahaux O, Bauchau V, Van Holle L. Pharmacoepidemiological considerations in observed-to-expected analyses for vaccines. *Pharmacoepidemiol Drug Saf*. 2016;25:215–22.

4 Gordillo-Marañón M, Candore G, Hedenmalm K, *et al.* Lessons learned on observed-to-expected analysis using spontaneous reports during mass vaccination. *Drug Saf*. Published Online First: 9 April 2024. doi: 10.1007/s40264-024-01422-8

5 Willame C, Dodd C, Gini R, *et al.* Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines. Zenodo 2021.

6 Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines (ACCESS-BGR). https://catalogues.ema.europa.eu/node/3578/administrative-details (accessed 25 April 2024)

7 Li X, Ostropolets A, Makadia R, *et al.* Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study. *BMJ*. 2021;373:n1435.

8 Ostropolets A, Li X, Makadia R, *et al.* Factors influencing background incidence rate calculation: Systematic empirical evaluation across an international network of observational databases. *Front Pharmacol*. 2022;13. doi: 10.3389/fphar.2022.814198

9 Willame C, Dodd C, Durán CE, *et al.* Background rates of 41 adverse events of special interest for COVID-19 vaccines in 10 European healthcare databases - an ACCESS cohort study. *Vaccine*. 2023;41:251–62.

10 Phillips A, Jiang Y, Walsh D, *et al.* Background rates of adverse events of special interest for COVID-19 vaccines: A multinational Global Vaccine Data Network (GVDN) analysis. *Vaccine*. 2023;41:6227–38.

11 Burn E, Li X, Kostka K, *et al.* Background rates of five thrombosis with thrombocytopenia syndromes of special interest for COVID-19 vaccine safety surveillance: Incidence between 2017 and 2019 and patient profiles from 38.6 million people in six European countries. *Pharmacoepidemiol Drug Saf*. 2022;31:495–510.

12 Vaccine Monitoring Platform. European Centre for Disease Prevention and Control. 2023. https://www.ecdc.europa.eu/en/about-ecdc/what-we-do/partners-and-networks/eu-institutions-and-agencies/vaccine-monitoring (accessed 29 February 2024)

13 IQVIA disease analyzer Germany. https://catalogues.ema.europa.eu/node/1021/administrative-details (accessed 5 December 2024)

14 Integrated primary care information (IPCI). https://catalogues.ema.europa.eu/data-source/42618 (accessed 5 December 2024)

15 The information system for research in primary care (SIDIAP). https://catalogues.ema.europa.eu/node/1019/administrative-details (accessed 5 December 2024)

16 Clinical practice research datalink (CPRD) GOLD. https://catalogues.ema.europa.eu/node/1026/administrative-details (accessed 5 December 2024)

17 Dernie F, Corby G, Robinson A, *et al.* Standardised and reproducible phenotyping using distributed analytics and tools in the Data Analysis and real World Interrogation Network (DARWIN EU). *Pharmacoepidemiol Drug Saf*. 2024;33:e70042.

18 Hripcsak G, Albers DJ. High-fidelity phenotyping: richness and freedom from bias. *J Am Med Inform Assoc*. 2018;25:289–94.

19 Raventós B, Català M, Du M, *et al.* IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. *Pharmacoepidemiol Drug Saf*. 2024;33. doi: 10.1002/pds.5717

20 Burn E, Li X, Kostka K, *et al.* Background rates of five thrombosis with thrombocytopenia syndromes of special interest for COVID-19 vaccine safety surveillance: Incidence between 2017 and 2019 and patient profiles from 38.6 million people in six European countries. *Pharmacoepidemiol Drug Saf*. 2022;31:495–510.

21 Moll K, Lufkin B, Fingar KR, *et al.* Background rates of adverse events of special interest for COVID-19 vaccine safety monitoring in the United States, 2019-2020. *Vaccine*. 2023;41:333–53.

22 Nasreen S, Calzavara A, Buchan SA, *et al.* Background incidence rates of adverse events of special interest related to COVID-19 vaccines in Ontario, Canada, 2015 to 2020, to inform COVID-19 vaccine safety surveillance. *Vaccine*. 2022;40:3305–12.

23 Cullen LA, Grange Z, Antal K, *et al.* COVID-19 vaccine safety in Scotland - background rates of adverse events of special interest. *Public Health*. 2023;224:1–7.

24 Pillsbury A, Phillips A, Deng L, *et al.* Background incidence rates of selected adverse events of special interest (AESI) to monitor the safety of COVID-19 vaccines. *Vaccine*. 2023;41:3422–8.

25 Lai FTT, Huang L, Chui CSL, *et al.* Multimorbidity and adverse events of special interest associated with Covid-19 vaccines in Hong Kong. *Nat Commun*. 2022;13:411.

26 Banga S, Khromava A, Serradell L, *et al.* Background incidence rates of health outcomes of interest for COVID-19 vaccine safety monitoring in a US population: a claims database analysis. *BMJ Open*. 2024;14:e083947.

27 Nyberg F, Lindh M, Vanfleteren LEGW, *et al.* Adverse events of special interest for COVID-19 vaccines - background incidences vary by sex, age and time period and are affected by the pandemic. bioRxiv. 2021;2021.10.04.21263507.

28 Clothier HJ, Shetty AN, Mesfin Y, *et al.* What would have happened anyway? Population data source considerations when estimating background incident rates of adverse events following immunisation to inform vaccine safety. *Vaccine*. 2024;42:1108–15.

29 Russek M, Quinten C, de Jong VMT, *et al.* Assessing heterogeneity of electronic health-care databases: A case study of background incidence rates of venous thromboembolism. *Pharmacoepidemiol Drug Saf*. 2023;32:1032–48.

30 Pottegård A, Lund LC, Karlstad Ø, *et al.* Arterial events, venous thromboembolism, thrombocytopenia, and bleeding after vaccination with Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: population based cohort study. *BMJ*. 2021;373:n1114.

31 Hormazábal-Aguayo I, Ezzatvar Y, Huerta-Uribe N, *et al.* Incidence of type 1 diabetes mellitus in children and adolescents under 20 years of age across 55 countries from 2000 to 2022: A systematic review with meta-analysis. *Diabetes Metab Res Rev*. 2024;40:e3749.

32 Astrup K, Olivarius N de F. Frequency of spontaneously occurring postmenopausal bleeding in the general population. *Acta Obstet Gynecol Scand*. 2004;83:203–7.

33 Jarach CM, Lugo A, Scala M, *et al.* Global prevalence and incidence of tinnitus: A systematic review and meta-analysis. *JAMA Neurol*. 2022;79:888–900.

34 Stohler NA, Reinau D, Jick SS, *et al.* A study on the epidemiology of tinnitus in the United Kingdom. *Clin Epidemiol*. 2019;11:855–71.

35 van den Boom L, Buchal G, Kaiser M, *et al.* Multimorbidity among adult outpatients with type 1 diabetes in Germany. *J Diabetes Sci Technol*. 2022;16:152–60.

17. ANNEXES

Appendix I: Cohort code use: the codes that led to inclusion in an AESI cohort.