Study Report

P4-C2-018

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

30/07/2026

Version 3.0

Authors: Ivan Lam, Daniel Prieto-Alhambra

Public

This document represents the views of the DARWIN EU® Coordination Centre only and cannot be interpreted as reflecting those of the European Medicines Agency or the European Medicines Regulatory Network.

CONTENTS

Study title[1]	DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe – supporting international preparedness
Study report version	V3.0
Date	30/07/2026
EUPAS number	EUPAS1000000969
Active substance	N/A
Medicinal product	N/A
Research question and objectives	Research question What are the background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe? Objectives The aim of this study was to estimate the background incidence rates of selected vaccine adverse events of special interest in Europe. The specific objectives of this study were: 1. To estimate population level incidence rates of selected adverse events of special interest (AESIs) in the general population between 2015 and the latest data availability, stratified by calendar period based on pre-, during-, and post-COVID-19 pandemic periods, calendar quarter, sex, age groups, sex and age groups, and data source 2. To estimate age-standardised incidence rates (to the European population) of the selected AESIs in the general population between 2015 and the latest data availability, stratified by calendar period based on pre-, during-, and post-COVID-19 pandemic periods, calendar quarter, sex, and data source 3. To compare the generated population-level background incidence rates across different healthcare settings and provide recommendations regarding the most suitable setting(s) to identify each AESI.
Countries of study	Denmark, Finland, Germany, The Netherlands, Norway, Spain, Sweden, The United Kingdom
Authors	Ivan Lam (i.lam@darwin-eu.org) Daniel Prieto-Alhambra (d.prietoalhambra@darwin-eu.org)

[1]This is a routinely repeated study from P3-C3-001 with EUPAS1000000254

LIST OF ABBREVIATIONS

Acronyms/terms	Description
ACCESS	vACCine covid-19 monitoring readinESS
AESI	Adverse events of special interest
ATC	Anatomical Therapeutic Chemical
BGR	Background rates
CC	Coordination Centre
CDM	Common Data Model
CNODES	Canadian Network for Observational Drug Effects Studies
COVID-19	Coronavirus disease-2019
CPRD GOLD	Clinical Practice Research Datalink GOLD
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
DRE	Digital Research Environment
EHR	Electronic Health Records
EMA	The European Medicines Agency
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
FinOMOP-THL	Finnish Care Register for Health Care
GBS	Guillain-Barré syndrome
GDPR	General Data Protection Regulation
GP	General Practitioner
GLMM	Poisson Generalised Linear Mixed Model
GVDN	Global Vaccine Data Network
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage
HSA	Health Sciences Authority
ICMRA	International Coalition of Medicines Regulatory Authorities
InGef RDB	InGef Research Database
IP	Inpatient
IPCI	Integrated Primary Care Information
LOINC	Logical Observation Identifiers Names and Codes
N/A	Not applicable
NLHR	Norwegian Linked Health Registry data
OHDSI	The Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
OT	Other

Acronyms/terms	Description
PHE	Public Health Emergencies
RWD	Real-World Data
RWE	Real-World Evidence
SIDIAP	The Information System for Research in Primary Care
SNOMED	Systematized Nomenclature of Medicine
WG	Working Group
WHO	World Health Organization

1. TITLE

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

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Ivan Lam Daniel Prieto Alhambra	University of Oxford University of Oxford, Erasmus MC
Data Scientist	Nuria Mercade Besora Edward Burn	University of Oxford University of Oxford
Clinical Domain Expert	Daniel Prieto Alhambra	University of Oxford, Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	**Names**	**Data Partner Organisation***
Danish Data Health Registries (DK-DHR)	Elvira Bräuner Susanne Bruun	Danish Medicines Agency
Finnish Care Register for Health Care (FinOMOP-THL)	Toni Lehtonen Laura Salonen	Finnish Institute for Health and Welfare (THL)
InGef Research Database (InGef RDB)	Raeleesha Norris Alexander Harms Annika Vivirito	InGef - Institute for Applied Health Research Berlin GmbH
Integrated Primary Care Information (IPCI)	Katia Verhamme Mees Mosseveld Marcel de Wilde Tineke de Ben	Erasmus MC
Norwegian Linked Health Registry data (NLHR)	Saeed Hayati Petrica Ioan Olteanu Hedvig Marie Egeland Nordeng	University of Oslo
Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)	Marcel Ballin Fredrik Nyberg Rickard Ljung Huiqi Li Talbäck Mats	Swedish Medical Products Agency - Gothenburg University
The Information System for Research in Primary Care (SIDIAP)	Elena Roel Herranz Augustina Giuliodori Picco Laura Granés Talita Duarte Salles	IDIAPJGol
Clinical Practice Research Datalink GOLD (CPRD GOLD)	Antonella Delmestri	University of Oxford

*Data partners do not have an investigator role. Data partners execute code at their data source, review, and approve their results.

3. ABSTRACT

Title

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

Rationale and background

The COVID-19 pandemic underscored the regulatory and public health need for timely post-authorisation safety monitoring of vaccines. Background incidence rates of adverse events of special interest (AESIs) enable observed-to-expected analyses, an essential first step to support signal management. Due to variability across countries, settings, data sources, and temporality, updated evidence is needed to support preparedness and regulatory decision-making.

This study aimed to support a proof-of-concept within the International Coalition of Medicines Regulatory Authorities (ICMRA) Working Group on Real-World Evidence for Public Health Emergencies, with the goal of generating background rates of AESIs selected based on capacity and interest at national level, and to discuss lessons learned of such studies. The present study, building on the previous DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe, supports vaccine safety monitoring at European level as part of the Vaccine Monitoring Platform.

Objectives

1. To estimate population level incidence rates of selected adverse events of special interest (AESIs) in the general population between 2015 and the latest data availability, stratified by calendar period based on pre-, during-, and post-COVID-19 pandemic periods, calendar quarter, sex, age groups, sex and age groups, and data source
2. To estimate age-standardised incidence rates (to the European population) of the selected AESIs in the general population between 2015 and the latest data availability, stratified by calendar period based on pre-, during-, and post-COVID-19 pandemic periods, calendar quarter, sex, and data source
3. To compare the generated population-level background incidence rates across different healthcare settings and provide recommendations regarding the most suitable setting(s) to identify each AESI.

Methods

Study design

Population-based cohort study

Population

The study population included all individuals observed in each of the participating data sources during the study period. We required individuals to have at least 365 days of data availability before entering the cohort. The index date of cohort entry was 1st January 2015 or the date that individuals fulfilled the data availability. The individuals must not have had a prior record of that specific outcome within a 90-day 'clean window' period.

Variables

Outcomes: Guillain-Barré Syndrome (GBS), myocarditis (+/- pericarditis), encephalitis, stroke (ischaemic stroke and haemorrhagic stroke), thrombocytopenia, and immune thrombocytopenia.

Relevant covariates: sex (female, male), age groups (0 – 4, 5 – 9, 10 – 19, 20 – 29, 30 – 39, 40 – 49, 50 – 59, 60 – 69, 70 – 79, 80 and over), calendar quarter (January – March, April – June, July – September, October –

December), and period based on pre- (2015 – 2019), during (2020 – 2022), and post-COVID-19 pandemic (2023 to latest available data)

Data sources

The study was conducted in eight data sources from eight European countries. Four, DK-DHR, FinOMOP-THL, NLHR, and HI-SPEED, consist of nationwide data source (primary/secondary care) data. One data source (InGef RDB) consists of health claims. Two data sources (IPCI and CPRD GOLD) consist of general practitioner Electronic Health Records (EHR) and one (SIDIAP) of primary care linked to hospital records.

Statistical analysis

Incidence rates per 100,000 person-years and corresponding 95% confidence intervals were reported. All analyses were stratified by calendar time based on pre-, during-, and post-COVID-19 pandemic periods, calendar quarter, age group, and sex within each data source. Incidence rates were not estimated if there were less than 5 events in each stratum. We also standardised these rates to the European population based on the European Standard Population 2013 for Objective 2. Meta-analyses of incidence rates of AESIs were conducted (with 95% prediction intervals) to pool estimates from data sources grouped according to healthcare setting.

Results

The rarest AESI was myocarditis with pericarditis, with background rates ranging from 0.09 (0.06 – 0.13) in CPRD GOLD to 2.91 (2.76 – 3.07) per 100,000 person-years in NLHR across the included data sources. Conversely, the highest rates were observed for thrombocytopenia, defined based on low platelet counts where laboratory measurements were available (DK-DHR, SIDIAP, and CPRD GOLD), where background rates ranged from 69.67 (68.96 – 70.38) in FinOMOP-THL and 2,548.91 (2,544.78 – 2,553.03) per 100,000 person-years in SIDIAP.

Crude incidence rates increased with age for most AESIs. A slight increase in incidence was however observed in young children aged 0–9 years (compared to those aged 10–19) for GBS, immune thrombocytopaenia, and encephalitis. Sex-stratified analyses showed higher incidence in males for several outcomes including GBS, thrombocytopenia, myocarditis with and without pericarditis, and stroke.

Temporal trend analyses indicated largely stable incidence over time, with evidence of seasonal variation in several AESIs. A transient reduction in incidence of some AESIs, including stroke and encephalitis, was observed during the early COVID-19 pandemic, followed by recovery to baseline levels post-pandemic. In contrast, an increase in myocarditis incidence was observed over the pandemic period across multiple data sources, dropping in the post-pandemic years.

Age-standardised incidence rates were largely consistent with crude incidence estimates across all AESIs.

Pooled incidence rates by healthcare settings showed higher incidence of all AESIs in nationwide data compared to general practitioner EHR, except for thrombocytopenia, which showed higher rates in general practitioner EHR. Low counts precluded accurate estimation of pooled rates for the rarest AESI.

Discussion

This study estimated background incidence rates of selected vaccine AESIs stratified by age, sex, and quarter across eight European data sources and different healthcare settings and proposes a scientifically rigorous framework to support the rapid and reproducible estimation of background rates of AESIs. The use of data mapped to Observational Medical Outcomes Partnership (OMOP) common data model and of reusable computable phenotypes has been key for the speed and scalability of these analyses and can support similar international efforts beyond Europe.

4. AMENDMENTS AND UPDATES

Amendment number	Previous approved version of the report	Date	Section of the study report	Amendment or update	Reason
1	V1.0	22/06/2026	9.1.5. Pooled crude incidence rates of AESIs by data source healthcare setting (Objective 3)	Pooled incidence rates reported were updated	A different statistical model was used to generate pooled estimates after communication with the EMA due to challenges with the estimation of prediction intervals in the presence of very low counts for some AESIs
2	V2.0	30/07/2026	9.1.5. Pooled crude incidence rates of AESIs by data source healthcare setting (Objective 3)	The labels for the data source groups used for pooled incidence rate estimates were revised	To better reflect the healthcare settings represented by the data sources grouped for meta-analysis

5. MILESTONES

Study deliverable	Timelines (planned*)	Timelines (actual)
Final Study Protocol	January 2026	February 2026
Creation of Analytical code	February 2026	March 2026
Execution of Analytical Code on the data	March 2026	March 2026
Draft Study Report	April 2026	April 2026
Final Study Report	May 2026	June 2026

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

6. RATIONALE AND BACKGROUND

The Coronavirus disease-2019 (COVID-19) pandemic highlighted the regulatory and public health need for timely post-authorisation safety monitoring of vaccines based on novel platforms (such as mRNA technology), but also for continuous monitoring throughout the lifecycle of established vaccines. Rapid regulatory response to a vaccine safety concern is crucial for maintaining public confidence in vaccination programs. Information on background incidence rates of adverse events of special interest (AESIs) may support the rapid response to a vaccine safety concern.[1] Observed-to-expected analyses serve as the essential first step to inform further signal strengthening and evaluation.[2-4] There is a continuous need for up-to-date evidence on background rates (BGRs) of AESIs for vaccines. Timely availability of reliable BGRs is essential for preparedness, both in early stages of vaccine safety signal evaluation (e.g. when introducing novel vaccines into vaccination campaigns) and for the monitoring of vaccine safety along the product lifecycle.

Previous studies have shown that background rates vary across age groups and sex and are often heterogeneous between regions and data sources.[5-7] This heterogeneity can originate from different clinical coding systems, health care delivery systems, clinical practice, or reflect true differences in the source population. It is therefore important to use the background rates that are generated from the same or similar data sources rather than rates estimated from different settings or data sources in observed-to-expected analysis. Understanding demographic and clinical characteristics of the population targeted for vaccination could provide useful information for evaluating vaccine safety signals in the future.[8] In addition, evidence from Real-World Data (RWD) studies published during and after the pandemic varies across countries and data sources, with variable granularity/stratification in terms of population groups and important factors such as seasonality. Therefore, updated evidence is needed to support preparedness and regulatory decision-making.

The Working Group on Real-World Evidence for Public Health Emergencies was established in 2024 by the International Coalition of Medicines Regulatory Authorities (ICMRA) to provide a forum for international regulators to proactively enhance the efficiency and coordination of responses to emerging public health emergencies (PHEs) through joint evidence generation. This initiative builds on the former COVID-19 Real-World Evidence (RWE) and Observational Studies Working Group,[9] addressing the need for collaboration beyond COVID-19 as emphasised by ICMRA in its 2022 statement on international collaboration to enable RWE for regulatory decision-making. The goal of the Working Group (WG) is to optimise preparedness by leveraging existing infrastructures across jurisdictions, to maintain 'ever-warm' governance and operational processes to support readiness when new PHEs arise.

This study aimed to support a proof-of-concept, with the goal of generating evidence on background rates of several selected vaccine AESIs within the WG and drawing lessons learned and recommendations for future studies. Two parallel studies with similar objectives are conducted by Health Canada using CNODES | Canadian Network for Observational Drug Effects Studies, and by the Health Sciences Authority (HSA) in Singapore, using national databases. A secondary goal of the WG is to critically review existing phenotypes to improve (where feasible) conceptual and operational definitions, and ultimately, to contextualise the findings from each jurisdiction to understand heterogeneity and its relevance for signal evaluation globally. This study also supports vaccine safety monitoring endeavours as part of the Vaccine Monitoring Platform.[10]

This study, building on the previous DARWIN EU® study (EUPAS1000000254) focused on estimating the background incidence rates of a shortlist of selected AESIs of regional interest for WG members, with greater granularity of data and refined phenotypes, in order to best support future observed-to-expected analyses for specific vaccine products.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

What are the background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe?

Objectives

The aim of this study was to estimate the background incidence rates of selected vaccine adverse events of special interest in Europe.

The specific objectives of this study were:

1. To estimate population level incidence rates of selected adverse events of special interest (AESIs) in the general population between 2015 and the latest data availability, stratified by calendar period based on pre-, during-, and post-COVID-19 pandemic periods, calendar quarter, sex, age groups, sex and age groups, and data source

2. To estimate age-standardised incidence rates (to the European population) of the selected AESIs in the general population between 2015 and the latest data availability, stratified by calendar period based on pre-, during-, and post-COVID-19 pandemic periods, calendar quarter, sex, and data source
3. To compare the generated population-level background incidence rates across different healthcare settings and provide recommendations regarding the most suitable setting(s) to identify each AESI.

8. RESEARCH METHODS

8.1. Study design

A cohort study was conducted using routinely collected health data from eight data sources from eight countries across Europe.

8.2. Follow-up

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

a) study start date (1st January 2015),

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

c) date on which they fulfilled the 'clean window' (details in **Section 8.6.1.**) 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 data source, or end of data availability of data source.

Individuals were allowed to re-enter the same outcome cohort more than once if they met the inclusion criteria of data availability and 'clean window'.

8.3. Study population with inclusion and exclusion criteria

General population:

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

8.4. Study setting and data sources

This study was conducted using routinely collected data from eight European data sources, including primary and secondary care data sources in the DARWIN EU® network. Four, namely DK-DHR, FinOMOP-THL, NLHR, and HI-SPEED, consist of nationwide data source (primary/secondary care) data. Two data sources, namely IPCI and CPRD GOLD, consist of general practitioner Electronic Health Records (EHR) data only. One data source (SIDIAP) consists of primary care linked to hospital records. One data source (InGef RDB) consists of health claims data.

All data were a priori mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). (**Table 1**)

Table 1. Data sources.

Country	Name of Data source	Health Care setting[1]	Type of Data	Number of active individuals	Calendar period covered by each data source
Denmark	Danish Data Health Registries (DK-DHR)	Community pharmacies, secondary care – specialists, hospital IP care	Registries	5.9M	1st January 2015 to 4th November 2025
Finland	Finnish Care Register for Health Care (FinOMOP-THL)	Primary care, secondary care – specialists (ambulatory or hospital OP care), hospital IP care	EHR, Registries	5.7M	1st January 2015 to 9th October 2024
Germany	InGef Research Database (InGef RDB)	Primary care – GPs, community pharmacists, primary care specialists, secondary care – specialists, hospital inpatient care	Claims	7.6M	1st January 2016 to 5th November 2025
The Netherlands	Integrated Primary Care Information (IPCI)	Primary care	EHR	1.33M	1st January 2015 to 15th October 2025
Norway	Norwegian Linked Health Registry (NLHR)[2]	Primary care, secondary care – specialists (ambulatory or hospital OP care), hospital IP care	Registries	5.55M	1st January 2015 to 31st December 2023
Spain	The Information System for Research in Primary Care (SIDIAP)	Primary care data source + linkage to hospital data	EHR	5.9M	1st January 2015 to 5th November 2025
Sweden	Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)[3]	Primary care, secondary care – specialists (ambulatory or hospital OP care), hospital IP care	Registries	10.6M	1st January 2016 to 31st August 2025
The United Kingdom	Clinical Practice Research Datalink (CPRD) GOLD	Primary care	EHR	2.6M	1st January 2015 to 20th May 2025

IP = inpatient, OP = outpatient, OT = other, n/a = not applicable, EHR = Electronic Health Record
Primary and specialist care data (excluding hospital records) from NLHR are available from 2008 to 31st December 2023, Hospital care data (in- and out-patient data) are available from 1st January 2018 till 31st December 2023.

Data sources selection

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

8.5. Study period

The study period was from 1st January 2015 to the most recent data available for each contributing data source.

8.6. Variables

8.6.1. Outcome

AESIs:

The AESIs of interest were selected by the ICMRA study team, based on national capacity and interest, while aligning across jurisdictions. The phenotypes had been defined in the previous study following the Standard Operating Procedure under a dynamic workflow between the study team and the EMA. Certain phenotypes were refined to accommodate new data partners for this analysis. In principle, broad and narrow phenotype variations were considered: “Broad”, a broader definition of the outcome, which included codes that are 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. Incidence of myocarditis with pericarditis was defined as having diagnosis records for both conditions on the same index date, whilst incidence of myocarditis without pericarditis was defined as the absence of a diagnosis record of pericarditis on the same index day as the diagnosis of myocarditis.

The AESIs of interest included Guillain-Barré Syndrome (GBS), myocarditis (+/- pericarditis), encephalitis, stroke (ischaemic and haemorrhagic stroke), thrombocytopenia, and immune thrombocytopenia.

For all study outcomes, a 90-day clean window was applied to define incident outcomes. Each AESI was considered incident if there was no record of the same outcome during the preceding 90 days. An individual had the potential to contribute multiple events if there was a gap of at least 90 days between each eligible event.

The concept sets used for the definition of outcomes are described in **Annex III**.

8.6.2. Covariates, including confounders, effect modifiers, intercurrent events, 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

Given that some events were very rare, with counts below 5 in the 5- and 10-years age groups, broader age groups were used as well: 0 – 18, 19 – 64, 65+.

Calendar time

- By quarter (January – March, April – June, July – September, October – December)
- By calendar period based on pre-specified windows: pre- (2015 – 2019), during (2020 – 2022), and post-COVID-19 pandemic (2023 to latest available data)

8.7. Study size

The population base for this study was approximately 73.2M, estimated based on the "total number of patients" from all the included data sources.

8.8. Data transformation

Analyses were conducted separately for each data source. Before study initiation, test runs of the analyses were performed on a subset of the data sources, and quality control checks were performed. Once all the tests passed (see **Annex III**), the final study codes package was released in the version-controlled Study Repository for execution against all the participating data sources.

The data partners locally executed the analytics against the OMOP CDM in R Studio and reviewed and approved the, by default, aggregated results.

The study results of all data sources were checked, after which they were made available to the team, and the dissemination phase started. All results were locked and timestamped for reproducibility and transparency

8.9. Statistical methods

8.9.1. Main summary measures

The main summary measure for this study was the incidence rate of separate study outcomes.

8.9.2. Main statistical methods

Objectives 1 and 2

Incidence rates per 100,000 person-years were calculated as the number of incident cases divided by the total person-time at risk. The corresponding 95% confidence intervals were reported using exact Poisson model. A pre-defined clean window of 90 days was applied to each AESI, where an individual can re-enter the cohort after the clean window of the previous event has been fulfilled. (**Table 2**)

The incidence rates were calculated using the *IncidencePrevalence*[11] R package, developed by DARWIN EU®. Analyses were stratified as follows:

- By data source
- By calendar quarter (January – March, April – June, July – September, October – December)
- By calendar period based on pre- (2015 – 2019), during (2020 – 2022), and post-COVID-19 pandemic periods (2023 to latest available data)
- By age group stated in **Section 8.6**. A wider age was used if most of the subgroups had event counts less than 5.
- By sex
- By age group and sex

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

The incidence rates were reported as crude estimates for objective 1 and further standardised by age to the European population by the direct method, using the same age groups for objectives 2. Age-standardised incidence was estimated based on the age distribution of the European population projected for 2011 – 2030 reported in the European Standard Population 2013.[12]

Table 2. Objectives 1 and 2 analyses specification.

Hypothesis:	Not applicable, as this study reports descriptive incidence rate.
Exposure contrast:	Not applicable
Outcome:	AESIs as listed in **Section 8.6.1.**[1]
Analytic software:	R
Model:	We estimated incidence rates and 95% confidence interval. Overall, age group, and sex specific rates were reported when event numbers were larger than 5 within the strata.
Confounding adjustment method	
	For incidence rates, we estimated the rates with stratification and standardisation
Subgroup Analyses	
	Sex: male, female Age groups: 1) 0 – 4, 5 – 9, 10 – 19, 20 – 29, 30 – 39, 40 – 49, 50 – 59, 60 – 69, 70 – 79, 80 and over.

[1]AESI = Adverse events of special interest.

Objective 3

Meta-analyses of results based on a random effects Poisson Generalised Linear Mixed Model (GLMM) were conducted to synthesise pooled findings across data sources, stratified by the following healthcare settings and data types: 1) nationwide data (primary/secondary care): DK-DHR, FinOMOP-THL, NLHR, and HI-SPEED; and 2) general practitioner EHR data sources: IPCI and CPRD GOLD. Meta-analytic estimates of incidence of AESIs with 95% prediction intervals were computed.

8.9.3. Sensitivity analysis

Sensitivity analyses were conducted by 1) using a narrow (but more specific) case definition on selected outcomes and 2) extending the clean windows of outcome to 365 days. Detailed descriptions of sensitivity analyses are presented by means of **Table 3**.

Table 3. Sensitivity analyses – rationale, strengths, and limitations.

	What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Narrower case definition for outcome events	A phenotype variation, tailored to each event and clinically relevant, was defined during the phenotyping process, then applied in the sensitivity analysis. The outcomes for which a narrower phenotype variation	The narrower phenotype variation of the outcome impacted on the included study population and outcome cases. If the case definition was narrower, we expected to see lower incidence rates for the specific event.	A narrower phenotype variation increased the specificity of incident event	The narrower phenotype variation for outcome could result in under capturing of incident cases of specific outcome

	What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
	were applied included: Encephalitis, Stroke, Ischaemic stroke, and Immune thrombocytopenia			
Extended clean window	A longer clean window of 365 days for all outcomes was implemented to define incident cases of outcomes	A longer clean window impacted on the duration of the period during which individuals with history of a specific outcome may re-enter the cohort of at-risk population, we might see a lower number of cases and incidence rate for the specific event.	A longer clean window would increase the specificity of the inclusion of incident events	A longer clean window could lead to under capturing incident cases of a specific outcome

8.10. Deviations from the protocol

Deviation number	Protocol version	Date	Section of study protocol	Deviation	Reason
1	3.0	19/05/2026	8.8. Analysis	Analysis on sex standardised incidence rates was not performed	The European standard population for male and female reported were identical across each age-strata. Additional standardisation by sex is therefore not required

9. RESULTS

The full set of results for this study are available through an interactive web-application Shiny app at EUPAS1000000969. The Shiny app consists of four main tabs:

1. Background – Provides a brief description of the study, including the rationale, background, and objectives.
2. Data Source Summary– Includes a Snapshot subtab with a table summarising metadata for each study data source.
3. Cohort Summary – Displays selected patient-level demographic, available in table format.
4. Incidence – Tables and plot tabs presenting estimated incidence rates

9.1. Main results

9.1.1. Crude incidence rate of AESIs (Objective 1)

The crude overall incidence rates of AESIs ordered from higher (top) to lower rates (bottom) are presented in **Figure 1.**

The crude incidence rate of Guillain-Barré syndrome ranged from 1.87 (1.73 – 2.02) to 6.36 (6.15 – 6.57) per 100,000 person-years observed in CPRD GOLD and FinOMOP-THL, respectively.

Thrombocytopenia was the most common AESI, with incidence ranging between 69.67 (68.96 – 70.38) in FinOMOP-THL and 2,548.91 (2,544.78 – 2,553.03) per 100,000 person-years in SIDIAP. The incidence of

immune thrombocytopenia was similar for the two proposed phenotype variations (broad and narrow), as seen in Figure 1. We therefore report here rates for the broad version, which ranged between 6.40 (6.14 – 6.66) in CPRD GOLD and 31.00 (30.54 – 31.46) per 100,000 person-years in DK-DHR.

The crude incidence of myocarditis with pericarditis ranged between 0.09 (0.06 – 0.13) in CPRD GOLD and 2.91 (2.76 – 3.07) per 100,000 person-years in NLHR. No cases of myocarditis with pericarditis were observed in IPCI. Myocarditis without pericarditis was more common, with crude incidence rates ranging from 2.09 (1.84 – 2.37) in IPCI to 34.74 (34.20 – 35.28) per 100,000 person-years in NLHR.

The incidence of stroke and ischaemic stroke varied considerably between narrow and broad phenotype variations. We report here the rates for both the broad and narrow variation. The incidence rate of stroke (broad) ranged from 76.49 (75.58 – 77.40) in CPRD GOLD to 573.04 (571.02 – 575.06) in FinOMOP-THL, whilst the rate of stroke (narrow) ranged from 46.26 (45.55 – 46.97) in CPRD GOLD to 121.06 (120.13 – 121.99) in FinOMOP-THL. The incidence rate of ischaemic stroke (broad) ranged from 50.83 (50.09 – 51.58) in CPRD GOLD to 463.60 (461.78 – 465.42) in FinOMOP-THL, whilst the rate of ischaemic stroke (narrow) ranged from 0.20 (0.17 – 0.24) in DK-DHR to 20.16 (19.70 – 20.64) in CPRD GOLD. In the broad definition of both stroke and ischaemic stroke, the incidence in CPRD GOLD were considerably lower in comparison with other data sources.

The crude incidence of haemorrhagic stroke was lower than ischaemic, with rates ranging between 26.31 (25.78 – 26.85) in CPRD GOLD and 120.45 (119.53 – 121.38) per 100,000 person-years in FinOMOP-THL.

The crude incidence of encephalitis was largely consistent for the broad and narrow definitions (see Figure 1). We therefore report here the incidence rates for the broad variation, which ranged between 2.93 (2.76 – 3.12) in CPRD and 21.02 (20.64 – 21.42) per 100,000 person-years in FinOMOP-THL.

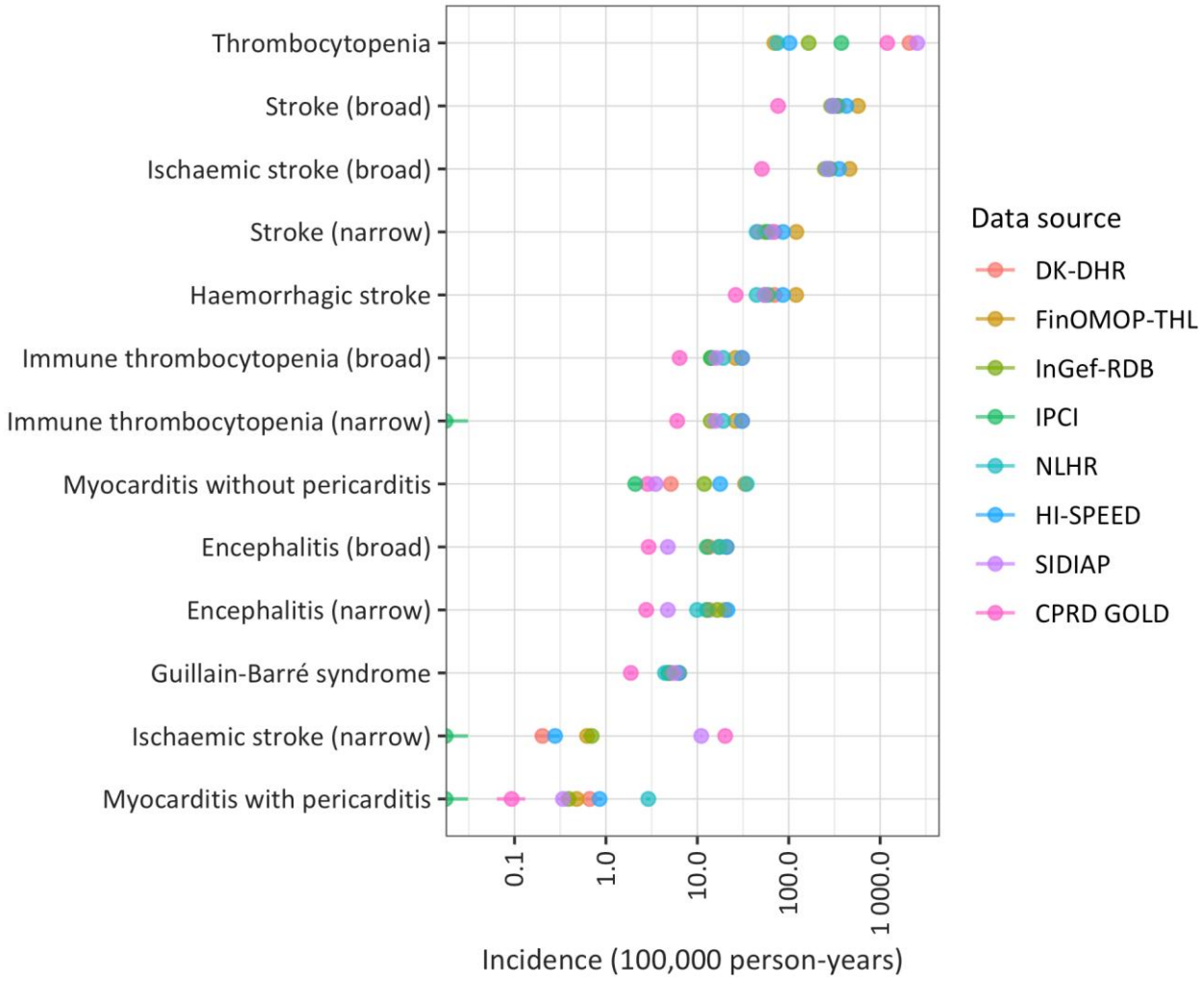

Figure 1. Crude incidence rates per outcome stratified by data source over the study period.

DK-DHR = Danish Data Health Registries, FinOMOP-THL = Finnish Care Register for Health Care, InGef RDB = Institute for Applied Health Research Berlin GmbH, IPCI = Integrated Primary Care Information, NLHR = Norwegian Linked Health Registry, SIDIAP = The Information System for Research in Primary Care, HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD = Clinical Practice Research Datalink GOLD

9.1.2. Crude incidence rate of AESIs stratified by age groups and sex

The age-stratified crude incidence rate of AESIs presented in **Figure 2** showed a monotonic association with age, with an increase in incidence rate with older age for most AESIs. Exceptions include immune thrombocytopenia and encephalitis, which show more of a J-shaped association, with high incidence rates in early childhood (0 – 4years old) that decline until around age 30 and then increase again with older age. Similarly, the background rate of myocarditis with pericarditis peaks in young adults (20 – 29 years old) only to then decrease in older ages. Myocarditis without pericarditis follows a similar pattern, but with a plateau in most data sources from age 20 to age 50, followed by a decline in older populations.

A higher crude incidence rate was observed in males for most AESIs, including: Guillain-Barré syndrome (GBS), thrombocytopenia, myocarditis with and without pericarditis, and both ischaemic and haemorrhagic stroke. The opposite was observed for immune thrombocytopenia (broad and narrow phenotypes) and encephalitis (broad and narrow), which appeared to have higher sex-specific rates in females versus males.

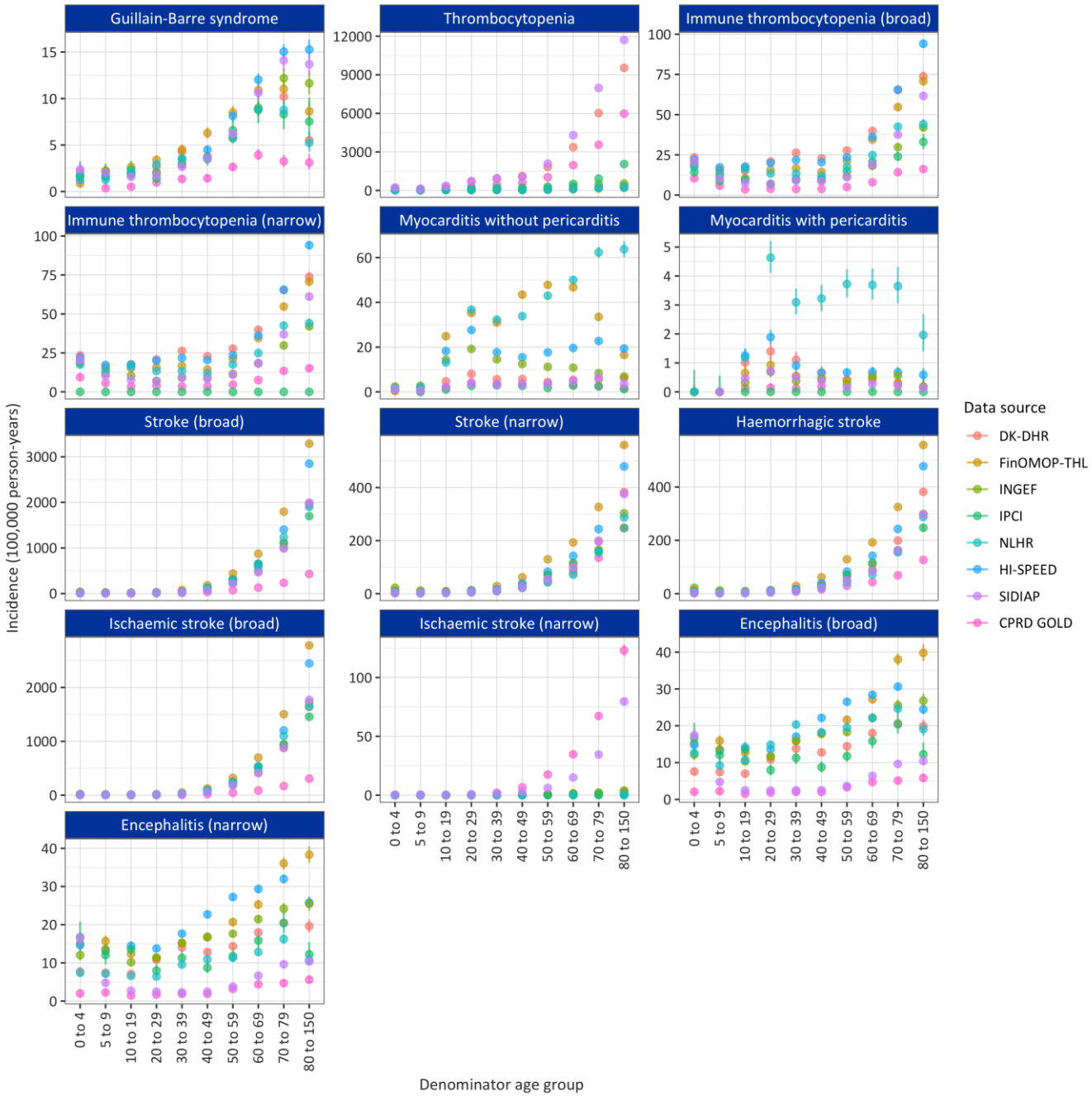

Figure 2. Crude incidence rates per outcome stratified by age groups.

DK-DHR = Danish Data Health Registries, FinOMOP-THL = Finnish Care Register for Health Care, InGef RDB = Institute for Applied Health Research Berlin GmbH, IPCI = Integrated Primary Care Information, NLHR = Norwegian Linked Health Registry, SIDIAP = The Information System for Research in Primary Care, HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD = Clinical Practice Research Datalink GOLD

9.1.3. Crude incidence rate of AESIs stratified by calendar quarter

The crude incidence rates of AESIs stratified by calendar quarter are presented in **Figure 3**. The crude quarter-specific incidence rate of GBS was largely consistent over time in most data sources, with a consistently observed increase in autumn – winter in NLHR, FinOMOP-THL, and SIDIAP. Compared to the pre-pandemic period, an increase in incidence of GBS during the post-COVID-19 pandemic period was observed in FinOMOP-THL and SIDIAP. In contrast, a reduction in incidence was observed in HI-SPEED over the same period.

A largely consistent incidence of thrombocytopenia and immune thrombocytopenia (broad and narrow) was observed in all data sources across the study period. A seasonal increase in incidence of thrombocytopenia was prominent in spring – summer in CPRD GOLD, DK-DHR, IPCI, and SIDIAP. A higher incidence of thrombocytopenia was observed over the post-COVID-19 pandemic period compared to the level observed in the pre-COVID-19 pandemic period.

The crude incidence of myocarditis without pericarditis showed a seasonal increase in incidence in autumn – winter compared to spring – summer. Across the study period, a greater incidence was also observed during the COVID-19 pandemic period in all data sources, which continued to persist over the post-COVID-19 period.

Seasonal variation in the crude incidence of ischaemic stroke was observed in all data sources, with consistent lower incidence observed in summer and higher incidence in winter. Overall, the crude incidence of ischaemic stroke (broad) was stable over time in most data sources, with an increase in incidence observed during the COVID-19 pandemic period in DK-DHR, NLHR, and SIDIAP. The temporal pattern for the crude incidence of haemorrhagic stroke was broadly similar to that observed for ischaemic stroke.

Crude incidence of encephalitis (broad and narrow) remained generally stable over the study period, with incidence observed over the pre-, during-, and post-COVID-19 pandemic period being largely consistent across all data sources. However, a progressive increase in incidence rate was observed over the COVID-19 pandemic period. A recurrent seasonal pattern with a rise in incidence during summer was observed across FinOMOP-THL, HI-SPEED, and InGef RDB.

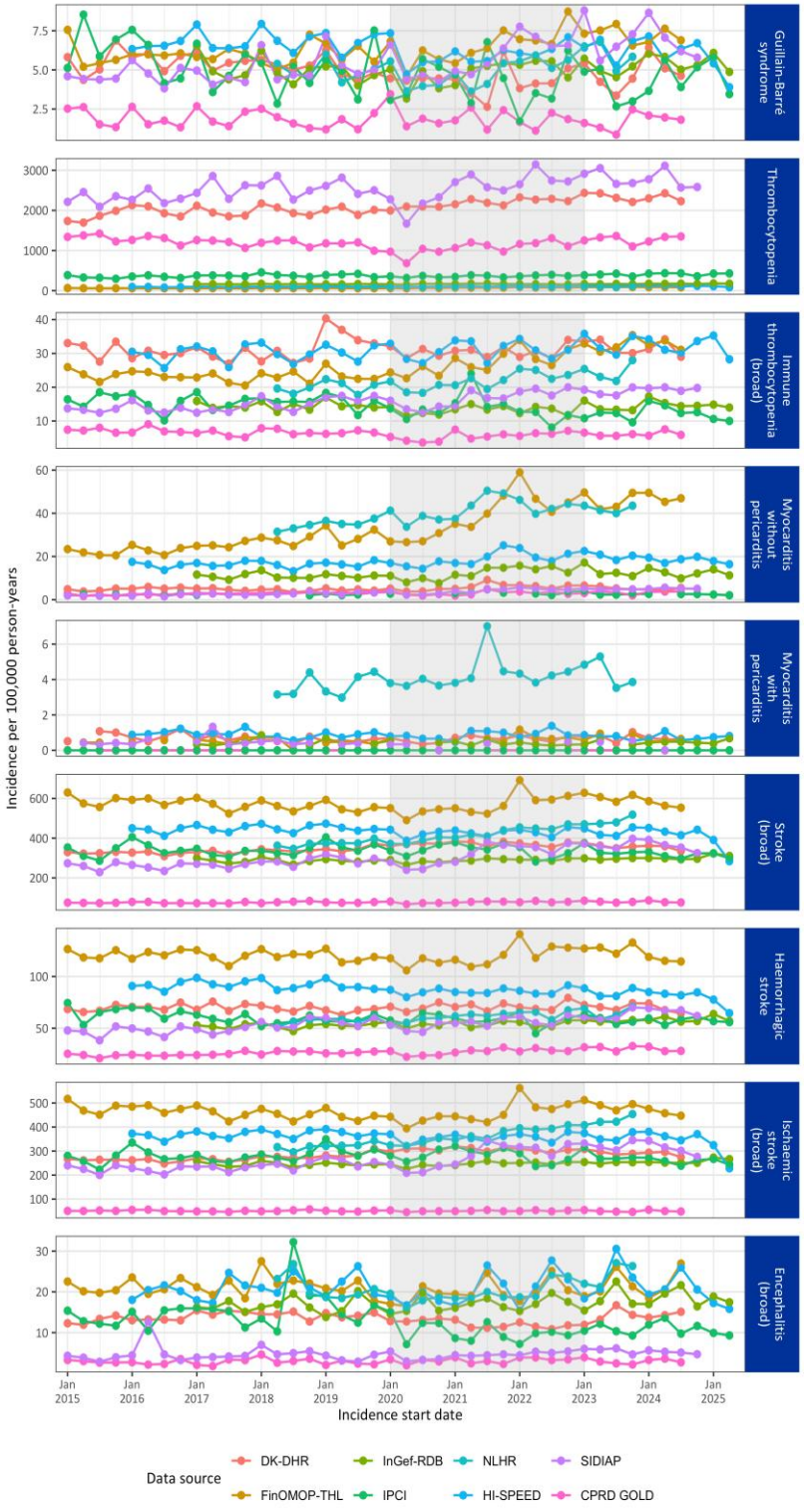

Figure 3. Crude incidence rates per outcome stratified by calendar quarter and year.

*The shaded regions denote the COVID-19 pandemic period (2020–2022).
*Incidence rate prior to 2018 were not presented in NLHR due to the lack of hospital record prior to this year.
DK-DHR = Danish Data Health Registries, FinOMOP-THL = Finnish Care Register for Health Care, InGef RDB = Institute for Applied Health Research Berlin GmbH, IPCI = Integrated Primary Care Information, NLHR = Norwegian Linked Health Registry, SIDIAP = The Information System for Research in Primary Care, HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD = Clinical Practice Research Datalink GOLD

9.1.4. Age-standardised incidence rates of AESIs (Objective 2)

The age-standardised incidence rates of AESIs ordered from higher (top) to lower rates (bottom) are presented in **Figure 4.**

The age-standardised incidence rates of GBS ranged from 1.85 (1.71 – 1.99) in CPRD GOLD to 6.22 (6.06 – 6.38) per 100,000 person-years in HI-SPEED.

The age-standardised incidence of thrombocytopenia ranged from 66.73 (66.05 – 67.40) in FinOMOP-THL to 2,492.19 (2,488.12 – 2,496.26) per 100,000 person-years in SIDIAP. The age-standardised incidence of immune thrombocytopenia was similar for the two proposed phenotype variations (broad and narrow). We therefore report here rates for the broad version, which ranged between 6.44 (6.18 – 6.71) in CPRD GOLD and 30.83 (30.37 – 31.29) per 100,000 person-years in DK-DHR.

The age-standardised incidence of myocarditis without pericarditis ranged between 2.10 (1.84 – 2.37) in IPCI and 35.50 (34.95 – 36.04) per 100,000 person-years in NLHR. Myocarditis with pericarditis was less frequent, with age-standardised incidence ranging between 0.09 (0.06 – 0.12) in CPRD GOLD and 2.91 (2.75 – 3.07) per 100,000 person-years in NLHR.

The incidence of stroke and ischaemic stroke varied considerably between narrow and broad phenotype variations. We report here the rates for both the broad and narrow variation. The incidence rate of stroke (broad) ranged from 75.29 (74.39 – 76.19) in CPRD GOLD to 528.86 (526.99 – 530.73) in FinOMOP-THL, whilst the rate of stroke (narrow) ranged from 45.51 (44.82 – 46.21) in CPRD GOLD to 113.14 (112.27 – 114.01) in FinOMOP-THL. The incidence rate of ischaemic stroke (broad) ranged from 50.03 (49.30 – 50.76) in CPRD GOLD to 426.01 (424.34 – 427.69) in FinOMOP-THL, whilst the rate of ischaemic stroke (narrow) ranged from 0.20 (0.16 – 0.23) in DK-DHR to 19.84 (19.38 – 20.30) in CPRD GOLD.

The age-standardised incidence of haemorrhagic stroke was lower than ischaemic, with rates ranging between 25.88 (25.36 – 26.41) in CPRD GOLD and 112.58 (111.71 – 113.45) per 100,000 person-years in FinOMOP-THL.

The age-standardised incidence of encephalitis was largely consistent for the broad and narrow. We therefore report here the incidence rates for the broad variation, which ranged between 2.91 (2.74 – 3.09) in CPRD and 20.89 (20.60 – 21.18) per 100,000 person-years in HI-SPEED.

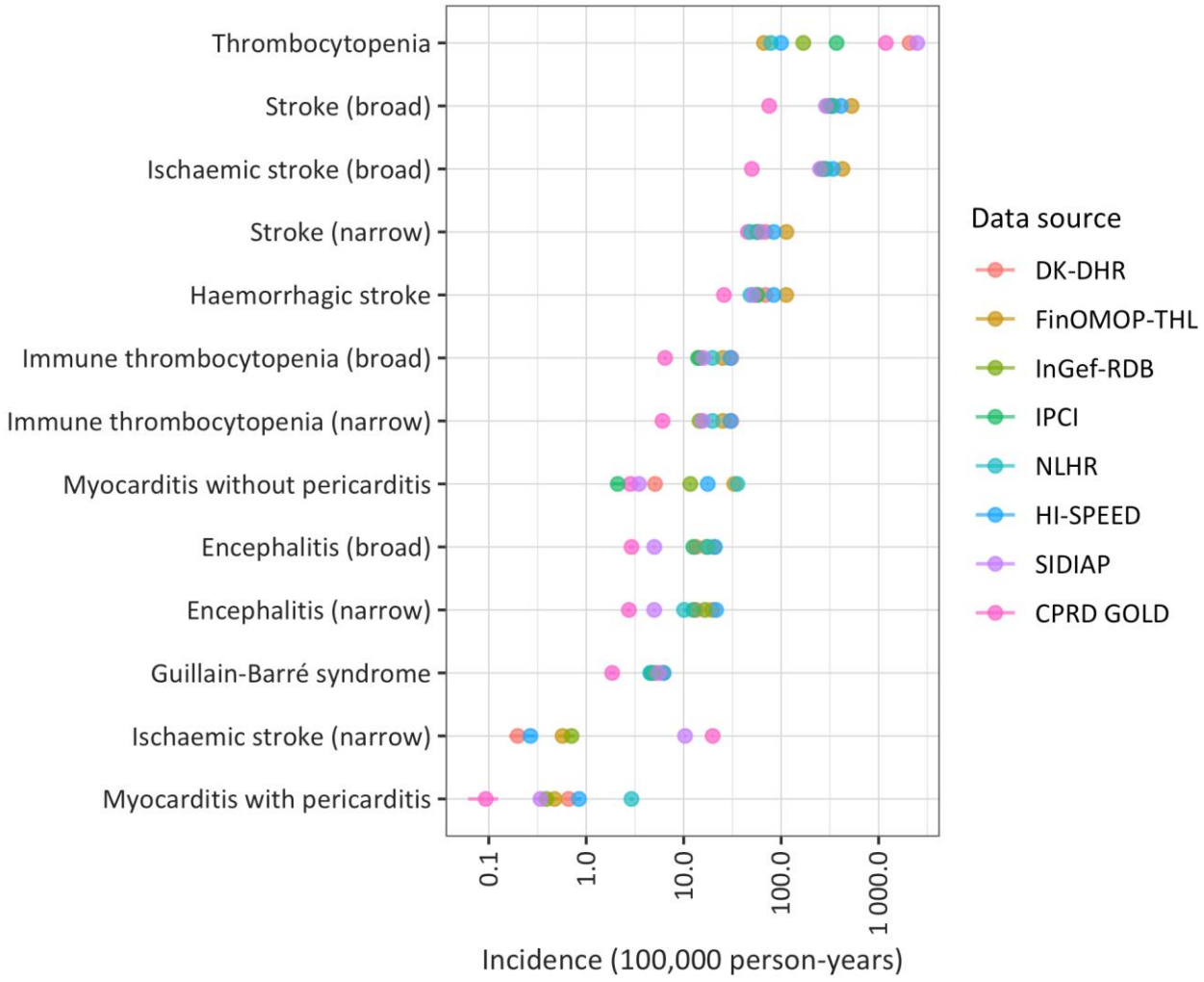

Figure 4. Age-standardised incidence rates per outcome stratified by data source over the study period.

DK-DHR = Danish Data Health Registries, FinOMOP-THL = Finnish Care Register for Health Care, InGef RDB = Institute for Applied Health Research Berlin GmbH, IPCI = Integrated Primary Care Information, NLHR = Norwegian Linked Health Registry, SIDIAP = The Information System for Research in Primary Care, HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD = Clinical Practice Research Datalink GOLD

9.1.5. Pooled crude incidence rates of AESIs by healthcare setting (Objective 3)

The crude incidence rates of AESIs grouped by healthcare settings were presented in **Figure 5**. Only incidence rates of broad phenotype variation were reported for AESIs with both broad and narrow phenotype variations.

A broadly higher incidence rate was observed in nationwide data sources for: immune thrombocytopenia, myocarditis without pericarditis, haemorrhagic, and ischaemic stroke.

A broadly similar incidence rate was observed in nationwide data sources for: GBS, myocarditis with pericarditis, and encephalitis (broad).

Pooled incidence rates of AESIs by healthcare settings reported overall higher incidence rates of most AESIs in nationwide data sources compared general practitioner EHR, except for thrombocytopenia. However, the prediction intervals for all outcomes were very wide due to low counts and therefore overlapped between the pooled estimates from general practitioner EHR and nationwide data sources data.

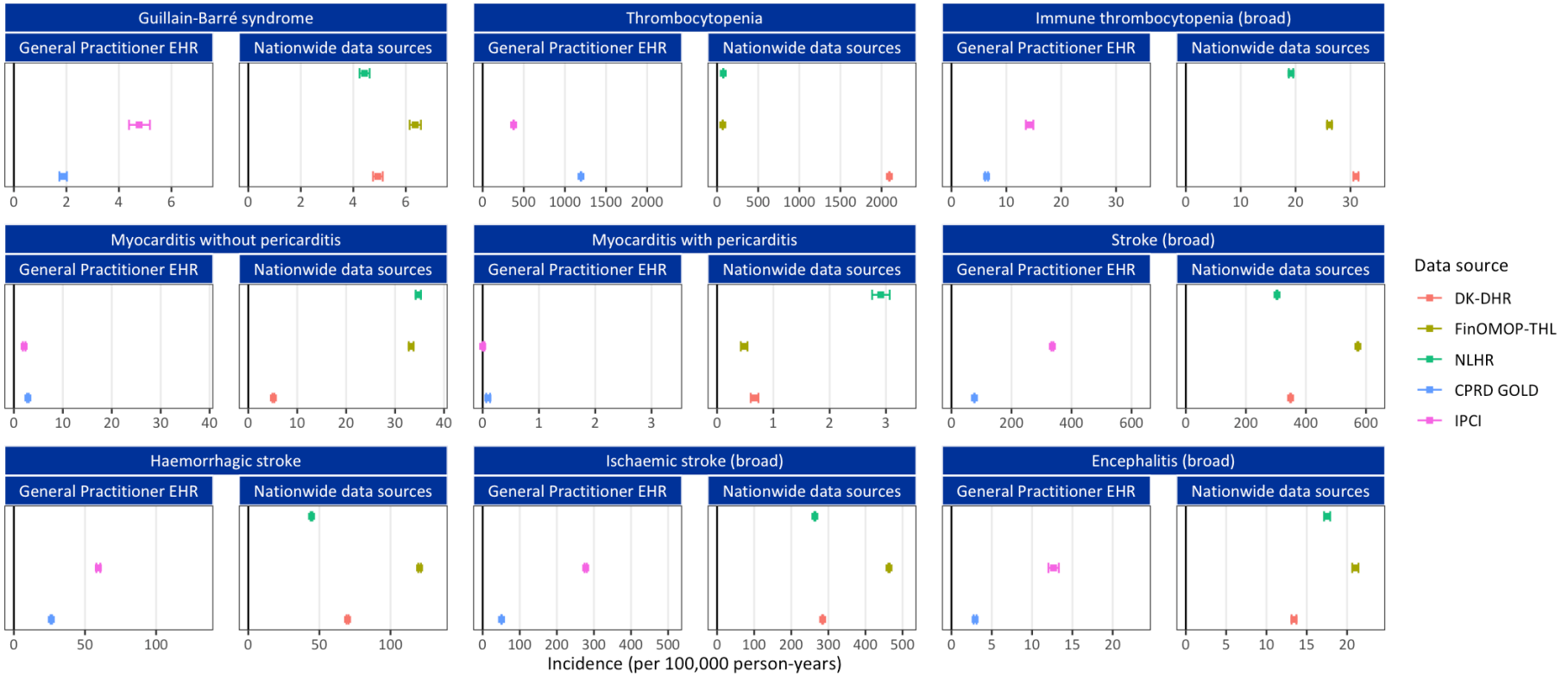

Figure 5. Crude incidence rates of AESIs in nationwide and general practitioner EHR data sources.

AESI = Adverse events of special interest, DK-DHR = Danish Data Health Registries, FinOMOP-THL = Finnish Care Register for Health Care, IPCI = Integrated Primary Care Information, NLHR = Norwegian Linked Health Registry, HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD = Clinical Practice Research Datalink GOLD, EHR = Electronic Health Records

9.2. Other analysis

Sensitivity analysis conducted by extending the clean windows of 365 days reported lower overall crude incidence rates for each of the study outcomes compared to those observed in the main analysis. The temporal trends in the crude incidence rate for each study outcome were largely consistent with those observed in the main analysis.

10. DISCUSSION

10.1. Key results

The rarest AESI was myocarditis with pericarditis, with background rates ranging from 0.09 in CPRD GOLD to 2.91 per 100,000 person-years in NLHR across the included data sources. Conversely, the highest rates were observed for thrombocytopenia, defined based on low platelet counts where laboratory measurements were available (DK-DHR, SIDIAP, and CPRD GOLD), where background rates ranged from 69.67 in FinOMOP-THL to 2,548.91 per 100,000 person-years in SIDIAP.

Crude incidence rates increased progressively with advancing age for the majority of AESIs, with higher rates consistently observed in older age groups compared to younger populations across all data sources. Conversely, some AESIs showed unique trends in their association with age. For example, myocarditis without pericarditis showed increased incidence rates amongst young adults aged 20 – 29, with a plateau in adults aged 40 – 69, whilst rates of myocarditis with pericarditis peaked among individuals aged 20 – 29 years old. A J-shaped association with age was observed for GBS, immune thrombocytopenia, and encephalitis, with high incidence rates in early childhood (<4 years old) which decline until around age 30, to then increase again with older age.

Regarding associations with sex, male individuals incur a higher incidence of GBS, thrombocytopenia, myocarditis with and without pericarditis, and both ischaemic and haemorrhagic stroke. Meanwhile, immune thrombocytopenia (broad and narrow) and encephalitis (broad and narrow) have a higher incidence in females.

Analysis of crude incidence rates stratified by calendar quarter revealed an increase in incidence of myocarditis with and without pericarditis during the COVID-19 pandemic period, whereas a mild reduction in incidence of stroke was observed during the same period.

Nationwide data sources reported higher incidence for most AESIs compared to general practitioner EHR , with the highest difference being observed for immune thrombocytopenia and myocarditis. One notable exception to this was thrombocytopenia, which showed comparable meta-analytic rates.

10.2. Strengths and limitations of the research methods

Strengths

The use of standardised methodologies across data sources improved comparability and reduced methodological heterogeneity, enabling more reliable interpretation of incidence patterns. The analyses were stratified by age and sex, quarterly calendar period, and before, during, and after the COVID-19 pandemic. These stratified analyses provided a reliable reference source on the background rates of separate AESIs taking into account the variability from the seasonal variation and the implications of COVID-19 pandemic of different AESIs. In general, comparisons of background incidence rate estimates across regions had been challenging due to differences in underlying population age composition. The use of age-standardised estimates in this study provided a more comparable assessment of incidence across different regions by minimising the influence of demographic variation between different populations. The phenotypes for all study outcomes were refined, critically reviewed and evaluated through diagnostics across participating data sources as part of the standard procedures established for DARWIN EU® studies to

ensure accurate cohort definitions in capturing cases of clinical outcomes across the different data sources.[14]

Furthermore, the standardised methodology and outcome phenotypes developed and implemented in this study provide a robust methodological framework for similar research initiatives in other regions. The use of harmonised methods enables direct comparison of incidence estimates across countries and healthcare settings, thereby supporting the development of a global evidence base on AESI background rates and strengthens preparedness for future public health emergencies, vaccine safety surveillance, and regulatory decision-making.

General limitations

The incidence rates estimated in this study only reflect populations from the included data sources. Electronic health records have certain inherent limitations because they are collected for clinical purposes rather than primarily for research use. It is assumed that if there are no related clinical codes of a condition presented for an individual in the data, the condition does not exist for this individual.

Study-specific limitations

The phenotypes implemented in this study may not be fully generalisable or directly applicable to later versions of the data sources or other data sources, requiring further diagnostics. Since published literature was not available for all AESIs to determine the appropriate length of the clean period, which could lead to misclassification of prevalent cases as incident, especially for outcomes with extended disease progression. However, sensitivity analyses conducted with a longer clean window of 365 days have reported findings consistent with the main analysis, suggesting that any potential misclassification of prevalent cases captured as incident cases should not have a major impact on the findings. Additionally, changes in clinical guidelines or recording practices could affect the estimation of incidence rates over time.

Under-capturing of outcomes is inherent if individuals received health care services outside of the data capture system. For example, events recorded in private care sectors may not be captured in the data sources included. However, the impact of this limitation on the observed incidence rates is expected to be minimal, given the comprehensive coverage of universal public healthcare systems in these countries. Outcome capture may also vary according to the healthcare setting represented by each data source. Data sources containing only primary care records, such as CPRD GOLD, under capture diagnoses made in secondary care settings. This is particularly relevant for serious acute conditions, including GBS and thrombocytopenia, which are typically diagnosed and managed in hospitals.

Abnormal platelet measurement records used to identify incidence of thrombocytopenia were unavailable in certain data sources, such as FinOMOP-THL, HI-SPEED, and InGef RDB. The inclusion of hospital health records from 2018 in NLHR could explain the substantial increase in the incidence observed for certain AESIs, such as stroke. The lower granularity of the source code used in Norwegian hospital source data for the common data model (CDM) mapping could result in potential underestimation of certain outcomes in NLHR.

10.3. Interpretation

Since the start of the COVID-19 pandemic, a substantial body of literature has focused on estimating background incidence rates of AESIs to support vaccine safety surveillance. Large multinational initiatives, including the vACCine covid-19 monitoring readinESS (ACCESS) project, the Global Vaccine Data Network (GVDN), and DARWIN EU®, have strengthened this evidence base by generating incidence estimates across multiple healthcare data sources using harmonised methodologies.[7, 15, 16]

Overall, the crude incidence rates observed in the present study were consistent with findings reported by the GVDN international surveillance studies. Similar to previous studies, substantial heterogeneity in incidence estimates was observed across data sources and regions, highlighting the influence of differences

in healthcare settings, coding practices, completeness of data capture, and underlying population characteristics. Thrombocytopenia was the most frequent AESI observed and was classified as common in accordance with the World Health Organisation (WHO) Council for International Organizations of Medical Sciences thresholds common in CPRD GOLD, DK-DHR, and SIDIAP.[13] The incidence rate estimates were considerably higher than that reported in the ACCESS study. Such differences could be attributed to the use of laboratory measurements to identify low platelet counts in our outcome definition. No laboratory measurements were used in the previous definition of thrombocytopaenia in ACCESS, leading to incomplete capture of cases of thrombocytopaenia.[7]

CPRD GOLD consistently demonstrated lower incidence estimates across multiple AESIs compared with other data sources. For certain outcomes, such as GBS, incidence estimates in CPRD GOLD were up to three-fold lower than those observed in other data sources. These findings suggest potential incomplete capture of cases within this data source, which may limit the representativeness of incidence estimates for other AESIs. In contrast, IPCI, and SIDIAP, which also comprise primary care data, reported incidence estimates that were consistent with those observed in other data sources, supporting representativeness.

Previous national and regional studies have identified a correlation between higher incidence estimates and advancing age as well as disparity by sex.[17-20] Accordingly, this study identified an increase in incidence with advancing age for most AESIs. Nonetheless, a J-shaped association with age, where a higher incidence was also observed in the infant and elderly populations, was also observed for certain AESIs. This finding was consistent with established epidemiological patterns and a multinational study based on healthcare records across the US, UK, Europe, and the Asia Pacific region.[5] The increased incidence of these outcomes observed among younger children may be related to immune immaturity in this age group, which could contribute to a greater risk of conditions, such as GBS, and encephalitis, which are frequently triggered by viral infections.[21, 22] Consistent with previous evidence, several AESIs, including stroke, and myocarditis with pericarditis, were observed more frequently among males than females.

Differences in seasonal patterns across outcomes likely reflect underlying variations in disease pathophysiology and the epidemiology of their triggers. For instance, the increase in encephalitis incidence observed in Northern European countries, including Finland, and Sweden, during the mid-year period, particularly between July and September, could be attributed to the seasonal activity of vector-borne infections, such as tick-borne encephalitis, which is highly prevalent in Europe in the summer and autumn between May and October.[23] In contrast, stroke has been shown to exhibit a winter peak and summer trough, with elder individuals appearing to be more susceptible to seasonal influences.[24] The increased incidence of GBS during winter periods observed in this study is consistent with previous studies supporting seasonal variation, with the effect reported to be more pronounced in Western countries compared to the Far and Middle East.[25] Meanwhile, the seasonal effects on the incidence of thrombocytopenia and myocarditis remained scarce. Further studies examining the seasonality of different AESIs are warranted to better understand these patterns and their underlying causes.

Given the heterogeneity in background incidence rates across patient demographics, healthcare settings, and calendar periods, it is important to reference contemporaneous estimates that are specific to the target population of interest. Consideration of temporal variation and seasonality is also essential when using background incidence rates to support the detection and interpretation of potential vaccine safety signals. Parallel studies conducted through the ICMRA collaboration could strengthen the global evidence base on AESIs background incidence rates by generating harmonised estimates across diverse healthcare systems, thereby supporting robust safety evaluations and future pandemic preparedness.

The age-standardised incidence rates reported in this study were consistent with age-standardised rates from the European ACCESS study, supporting the overall validity and representativeness of the findings across European healthcare populations.

The consistent crude and age-standardised estimates suggested that differences in overall incidence across data sources were not substantially influenced by variations in age distribution in the underlying populations, but more likely reflected differences in healthcare systems, data capture, coding practices, and case ascertainment.

Consistent with the pooled estimates by healthcare provenance reported in the ACCESS study, incidence rates across different healthcare settings could vary substantially, with estimates generated from primary care linked hospital data sources often higher compared to primary care only data sources. This study identified a higher incidence of most AESIs in nationwide data sources compared to general practitioner EHR, particularly on outcomes with severe clinical presentation which would typically require hospital admission such as stroke, myocarditis, and encephalitis. Nonetheless, the presence of laboratory measurement data (available in DK-DHR and all primary care data) have contributed to a higher incidence of thrombocytopenia in primary care only data sources compared to nation-wide registry data sources where laboratory measurements are unavailable.

The consistently lower rates observed in CPRD GOLD may be attributed to incomplete hospital diagnosis records. For instance, the incidence of GBS was up to three-fold lower in CPRD GOLD compared to other data sources. These findings suggest potential underestimation and limited representativeness of incidence rates for some AESIs in data sources where substantial disparities were observed relative to other data sources. In contrast, GBS incidence estimates from IPCI and SIDIAP, which also consist of mostly primary care data, were more aligned to the other data sources.

Conversely, the accurate diagnosis of thrombocytopenia requires blood test measurements, which are routinely conducted in primary care settings as part of the initial diagnosis. Data sources with more comprehensive laboratory measurement records, such as primary care data and DK-DHR, may therefore provide a more reliable estimate, while, the incidence is likely under-estimated in data sources without comprehensive laboratory results, as they primarily capture cases with recorded diagnoses, reflecting only more severe cases or cases requiring clinical attention.

The disparity observed between the pooled estimates of the broad and narrow ischaemic stroke phenotypes was primarily driven by the substantial reduction in cases captured by the narrow phenotype definition in registry-linked data sources and IPCI. In contrast, the narrower phenotype definition had a smaller impact on incidence estimates in CPRD GOLD, resulting in relatively higher pooled estimates among primary care data sources. The reduced incidence observed with the narrow phenotype was largely attributable to the exclusion of concepts related to cerebral infarction, cerebral thrombosis, occlusion of cerebral arteries, and lacunar infarction, which contributed to the considerably lower incidence observed in certain data sources.

The pooled estimates from the meta-analysis should be interpreted with caution, particularly for rare outcomes with low event counts and substantial heterogeneity across data sources, observed in primary care data sources. Random effect meta-analytic models can be sensitive to sparse data and between-data source variability, which may result in wide prediction intervals and, in some cases, implausible negative lower bounds. These negative values do not represent true negative incidence rates, but rather reflect statistical artefacts arising from the modelling of rare events in highly heterogeneous estimates.[26] Approaches such as Poisson GLMM allow to avoid negative lower bounds through the use of a log-link function and could produce more stable and interpretable estimates across outcomes. Nevertheless, prediction intervals will remain wide for rare outcomes with substantial between-data source heterogeneity, reflecting the underlying uncertainty on the pooled estimates.

Computable phenotyping

The phenotypes for each of the study outcomes were reviewed and refined using distributed analytics and tools, following the DARWIN EU® standard procedure. This supports more accurate capturing of specific

concepts used in new data sources in comparison to the previous study on background AESIs by DARWIN EU® and incorporation of concepts from the latest version of the OMOP CDM vocabulary.

A narrower definition, based on a highly specific set of concepts for selected outcomes was developed to assess the impact of using a more restrictive phenotype variation on the observed incidence rates. The incidence rate for outcomes defined by the broad and narrow definitions were largely consistent, apart from ischaemic stroke, where the incidence rate estimated with the narrow case definition was considerably lower across several data sources. Notably, concepts relating to cerebrovascular accident and neurological syndromes were removed from the definitions for ischaemic stroke, whilst concepts relating to cerebral and cerebellar infarction were further excluded from the narrow phenotype variation for ischaemic stroke. The main difference between narrow and broad definition is the inclusion of unspecific cerebral infarction codes such as READ G64z.00 (“Cerebral infarction NOS”) in CPRD GOLD or ICD-10 I63.9 (“Cerebral infarction, unspecified”) widely used in all included Scandinavian data and SIDIAP. These were included for the broad definition of ischaemic stroke, but not part of the definition of the narrow variation. Therefore, whether these codes are included should be carefully assessed in the study of the epidemiology of stroke in Real-World Data, as well as for the monitoring of this event in the context of pharmacovigilance and/or vaccine safety monitoring.

The absence of platelet measurement records in several data sources, including FinOMOP-THL, HI-SPEED, and NLHR, could contribute to the lower incidence of thrombocytopenia observed compared to other data sources from a similar healthcare setting. Conversely, the definition of immune thrombocytopenia did not rely on laboratory measurements, which may explain the greater consistency in incidence estimates across data sources.

The narrow definition of immune thrombocytopenia excluded concepts of thrombocytopenic purpura. However, as this is the only code used to identify cases of immune thrombocytopenia in IPCI, cases were not captured under this narrower definition in this data source.

10.4. Generalisability

The findings of this study provide updated estimates of background incidence rates for the selected AESIs, derived from large, population-based data sources across eight European countries. The inclusion of multiple data sources capturing diverse healthcare settings and populations enhances the generalisability of the results and allows for a comprehensive assessment of variability across regions, age groups, and data sources. The inclusion of population data from multiple countries strengthens the external validity of the estimates and ensures that they reflect real-world clinical practice rather than being specific to a single healthcare system.

The phenotypes implemented in this study were developed specifically for the data sources included and for the purpose of estimating background rates of AESIs of vaccines. To re-use these definitions in other data sources and/or for other research questions, 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 required.

11. CONCLUSION

This study estimated the background incidence rates of selected AESIs for vaccines stratified by age, sex, and quarter across eight European data sources across different healthcare settings. Through a standardised analytical approach, the findings from this study generated robust evidence to support future vaccine safety surveillance. This study demonstrates a scientifically rigorous framework to support the rapid and reproducible estimation of background rates of AESIs. The use of data mapped to a common data model (OMOP) and of reusable computable phenotypes has been key for the speed and scalability of these analyses and can support similar European and international efforts beyond Europe.

12. REFERENCES

1.Black SB, Law B, Chen RT, Dekker CL, Sturkenboom M, Huang W-T, et al. The critical role of background rates of possible adverse events in the assessment of COVID-19 vaccine safety. Vaccine. 2021;39(19):2712–8.

2.Mahaux O, Bauchau V, Van Holle L. Pharmacoepidemiological considerations in observed-to-expected analyses for vaccines. Pharmacoepidemiology and Drug Safety. 2016;25(2):215–22.

3.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. Pharmacoepidemiology and Drug Safety. 2023;32(7):783–94.

4.Gordillo-Marañón M, Candore G, Hedenmalm K, Browne K, Flynn R, Piccolo L, et al. Lessons Learned on Observed-to-Expected Analysis Using Spontaneous Reports During Mass Vaccination. Drug Safety. 2024;47(7):607–15.

5.Li X, Ostropolets A, Makadia R, Shoaibi A, Rao G, Sena AG, 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.

6.Ostropolets A, Li X, Makadia R, Rao G, Rijnbeek PR, Duarte-Salles T, et al. Factors Influencing Background Incidence Rate Calculation: Systematic Empirical Evaluation Across an International Network of Observational Databases. Frontiers in Pharmacology. 2022;Volume 13 - 2022.

7.Willame C, Dodd C, Durán CE, Elbers R, Gini R, Bartolini C, 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(1):251–62.

8.Burn E, Li X, Kostka K, Stewart HM, Reich C, Seager S, 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. Pharmacoepidemiology and Drug Safety. 2022;31(5):495–510.

9.Beck AE, Kampman M, Huynh C, Simon C, Plueschke K, Cohet C, et al. Collaborative Real-World Evidence Among Regulators: Lessons and Perspectives. Clinical Pharmacology & Therapeutics. 2025;117(2):368–73.

10. Vaccine Monitoring Platform: European Centre for Disease Prevention and Control; 2023 [Available from: https://www.ecdc.europa.eu/en/about-ecdc/what-we-do/partners-and-networks/eu-institutions-and-agencies/vaccine-monitoring.

11. Raventós B, Català M, Du M, Guo Y, Black A, Inberg G, et al. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. Pharmacoepidemiology and Drug Safety. 2024;33(1):e5717.

12. European Standard Population 2013 Public Health Scotland [updated 28 October 2025. Available from: https://publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/national-data-catalogue/national-datasets/search-the-datasets/european-standard-population/.

13. The Council for International Organizations of Medical Sciences (CIOMS). Guidelines for Preparing Core Clinical-Safety Information on Drugs Second Edition – Report of CIOMS Working Groups III and V. Geneva: 1999.

14. Dernie F, Corby G, Robinson A, Bezer J, Mercade-Besora N, Griffier R, et al. Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU). Pharmacoepidemiology and Drug Safety. 2024;33(11):e70042.

15. Phillips A, Jiang Y, Walsh D, Andrews N, Artama M, Clothier H, 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(42):6227–38.

16. DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe; EUPAS1000000254 [Available from: https://catalogues.ema.europa.eu/node/4155/administrative-details].

17. Banga S, Khromava A, Serradell L, Chabanon A-L, Pan C, Estevez I, 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(7):e083947.

18. Moll K, Lufkin B, Fingar KR, Ke Zhou C, Tworkoski E, Shi C, 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(2):333–53.

19. Nasreen S, Calzavara A, Buchan SA, Thampi N, Johnson C, Wilson SE, 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(24):3305–12.

20. Pillsbury A, Phillips A, Deng L, Quinn H, Macartney K, Gidding H. Background incidence rates of selected adverse events of special interest (AESI) to monitor the safety of COVID-19 vaccines. Vaccine. 2023;41(22):3422–8.

21. Duerlund LS, Nielsen H, Bodilsen J. Current epidemiology of infectious encephalitis: a narrative review. Clinical Microbiology and Infection. 2025;31(4):515–21.

22. Shahrizaila N, Lehmann HC, Kuwabara S. Guillain-Barré syndrome. The Lancet. 2021;397(10280):1214–28.

23. Beauté J, Spiteri G, Warns-Petit E, Zeller H. Tick-borne encephalitis in Europe, 2012 to 2016. Eurosurveillance. 2018;23(45):1800201.

24. Libruder C, Yaari R, Fluss R, Hershkovitz Y, Ram A, Tanne D, et al. Age-dependent seasonality in the incidence of stroke: A 21-year population-based study. European Stroke Journal. 2024;9(2):460–7.

25. Webb AJ, Brain SA, Wood R, Rinaldi S, Turner MR. Seasonal variation in Guillain-Barre syndrome: a systematic review, meta-analysis and Oxfordshire cohort study. J Neurol Neurosurg Psychiatry. 2015;86(11):1196–201.

26. Higgins JPT, Thompson SG, Spiegelhalter DJ. A Re-Evaluation of Random-Effects Meta-Analysis. Journal of the Royal Statistical Society Series A: Statistics in Society. 2009;172(1):137–59.

13. ANNEXES

ANNEX I. Description of data sources

Denmark: Danish Data Health Registries (DK-DHR)

#	Section	Description
1	Database Identification and country	DK-DHR (Danish Data Health Registries) Denmark
2	Data partner information section	Danish Medicines Agency (DKMA) Data Analytics Centre (DAC)
3	Coverage and timespan	Data collection since: 1995 Extent: Nationwide. The data is representative of the entire Danish population.
4	Healthcare setting / type of data	Community pharmacists, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: diagnosis (including rare diseases and pregnancy data), hospital admissions, discharge and ICU data, Cause of death, Drug prescriptions, dispensing, vaccination and contraception, Procedures, Devices, and Sociodemographic information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. All causes of deaths, all retrieved drug prescriptions, all records of vaccinations, all hospital inpatient and outpatients contacts including disease diagnoses and hospital surgical and non-surgical procedures, cancers, laboratory test results for the entire Danish population from 1/1/1995 onwards.
6	General representativeness	The data is representative of the entire Danish population. Healthcare is free in Denmark, so we do not expect any bias in data collection based on socio-economic status.
7	Data content /source coding	Diagnoses and causes of death are collected using the ICD-10 vocabulary. ATC and RxNorm are used for Drugs. SNOMED codes are used for Procedures.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network.
9	Quality control (database specific)	The data we have received relating to nationwide Danish Health Data registries offer an opportunity for large-scale, population-based studies with several advantages 1) Their large size improves the precision of estimates and enables the study of rare exposures and outcomes with long-term latency, 2) Inclusion of nearly all individuals in the target population ensures that the data reflect routine clinical care and all clinical segments of the source population, 3) Data are collected independently of each research study, thus minimising certain types of bias, e.g., non-response, and the influence from attention to the research question on the diagnostic process. Before the source data is sent to us, the Danish Health Data Authority runs and does comprehensive checks of the registry table data validity of the variables, breaks in data, changes in variable coding, missingness, etc. We perform checks of missingness/completeness in relation to requested variables. In essence, we are receiving a dump of a mirror of the data that is controlled by the SDS. The documentation performed by SDS is available online, in Danish primarily https://www.esundhed.dk/Dokumentation (all variables), but also in English https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/national_health_registers
10	Linkage	There is no linkage in this data source.
11	Vital status	The Cause of Death registry (DAR) is used, the cause of death is collected using ICD-10 codes.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Schmidt M, Schmidt SAJ,Adelborg K,Sundbøll J,Laugesen K,Ehrenstein V,Sørensen HT "The Danish health care system and epidemiological research: from health care contacts to database records." Clinical epidemiology (2019): 31372058

#	Section	Description
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111217 Website: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/healthdatadenmark

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

#	Section	Description
1	Database Identification and country	FinOMOP-THL (Finnish Care Register for Health Care) Finland
2	Data partner information section	Finnish Institute for Health and Welfare (THL) Department of Knowledge Brokers
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. The current CDM population comprises all persons having been alive and residing in Finland since the beginning of 2011.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The THL database covers both public and private, primary, and specialised inpatient and outpatient health care encounters in Finland, starting from 2011. The entire public sector and private inpatient encounters have been included since 2011, while private outpatient encounters, including occupational care, are included since 2020. Since 1998, the register has covered both public outpatient and inpatient specialized care and private inpatient care (TerveysHilmo). Since 2009, the Finnish National Vaccination Register is covered (complete since 2020). The vaccination register covers all vaccinations from the public sector and from a large part of private vaccination providers, with the data coverage from both sections being very good from 2020 onwards. Since 2011, the register has covered public primary care (AvoHilmo). Since 2020, the register has covered private outpatient care and occupational care. In addition, the CDM also contains positive COVID-19 test results from the Finnish National Infectious Diseases Register, which is maintained by THL.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. Data is entered by clinicians upon healthcare contact and processed by THL.
6	General representativeness	The THL data has national coverage and is therefore well representative of the Finnish population. Using the complete population as a basis for the person table also serves to facilitate calculations on a population level, e.g. incidence rates.
7	Data content /source coding	The following coding systems have been OMOP-mapped, typically to a good level of completeness: ICD10fi Finnish Extension, ATC, Toimenpideluokitus (procedure classification adapted from the Nordic Classification of Surgical Procedures (NCSP)), Terveydenhuollon erikoisalat (Hilmo specific provider speciality), Rokotustapa (AR/YDIN National classification for vaccine administration), Tupakointistatus (AR/YDIN National classification for smoking status). Vaccinations are identified on product level based on batch number, trade name, vaccine title, and ATC-code. This is mapped on brand and type in the OMOP CDM.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Each patient in THL has a unique identifier.
9	Quality control (database specific)	The source data collection undergoes a structural and semantic validation before entry into the source database. Additionally, some coded variables undergo quality assessment against the respective code systems post entry into the database. The source registers are also assessed for completeness and coverage, with the aim of improving future collection in the areas where data is lacking.

#	Section	Description
10	Linkage	THL is already a linkage of multiple Finnish registries (see above).
11	Vital status	The National Population registry data forms the basis for forming the patient population. This ensures an up-to-date location (municipality of residence) of patients, as well as complete death occurrences (although not the cause of death).
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Häkkinen, Pirjo; Mölläri, Kaisa; Saukkonen, Sanna-Mari; Väyrynen, Riikka; Mielikäinen, Lasse; Järvelin, Jutta "Hilmo – Sosiaali- ja terveydenhuollon hoitoilmoitus 2020 : Määrittelyt ja ohjeistus : Voimassa 1.1.2020 alkaen" Terveyden ja hyvinvoinnin laitos (2019):
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111187 Website: https://thl.fi/fi/tilastot-ja-data/ohjeet-tietojen-toimittamiseen/hoitoilmoitusjarjestelma-hilmo

Germany: InGef Research Database (InGef RDB)

#	Section	Description
1	Database Identification and country	InGef RDB (InGef Research Database) Germany
2	Data partner information section	Institut für angewandte Gesundheitsforschung Berlin GmbH
3	Coverage and timespan	Data collection since: 2014 Extent: Nation-wide. The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDC currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs. The data in the database depicts all health care use which has been reimbursed by the SHI.
4	Healthcare setting / type of data	Primary care – gps, and community pharmacists, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data. The following data elements are collected: pregnancy data, hospital admission, and/or discharge also with ICU admission. Prescription, dispensing drugs, and Advanced therapy medicinal products. Contraception, medical devices, vaccinations, procedures, diagnoses, and demographic information.
5	Data collection process	Insurance/administrative claims. The data in the database depicts all health care use which has been reimbursed by the SHI (statutory health insurances).
6	General representativeness	The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDC currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs.
7	Data content /source coding	The ATC and OPS (Operationen- und Prozedurenschlüssel) are used for prescription and dispensing drugs. For Procedures, the EBM (Einheitlicher Bewertungsmaßstab – doctor's fee scale) and for ambulatory procedures; OPS (Operationen- und Prozedurenschlüssel) for operations conducted at the hospital are used. Medical events are coded in ICD-10-GM and another vocabulary used is PZN (Pharmazentralzummer -pharmaceutical reference number).
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. No. In the German statutory health system, a person can only be enrolled in one health insurance at a

#	Section	Description
		time. However, if a person changes from one contributing insurer to another, a new ID number will be generated.
9	Quality control (database specific)	Before entering the InGef database, the data elements are checked with respect to data format, completeness, and plausibility. After each data update, data are compared with the previous data update in regard to number of records, number of data providers, etc. Due to the anonymized nature of the database, no direct validation of the data (e.g. using medical charts as the gold standard) is possible. Data delivery by health care providers is generally based upon standardized data requirements and formats provided by the National Association of Statutory Health Insurance Funds (compare: https://www.gkv-datenaustausch.de/leistungserbringer/leistungserbringer.jsp)
10	Linkage	No
11	Vital status	The Cause of Death is not captured, the date of death is captured.
12	Limitations	Ambulatory diagnosis are received from the source on a quarterly basis. These diagnoses are mapped to the observation table with the concept_id History of event within 3 months (1340222), with the actual diagnosis concept_id recorded in the field value_as_concept_id, and the date as the last day of the respective quarter (i.e. 30/31st of Mar/Jun/Sept/Dec). Ambulatory prescriptions are available with exact dates. The cause of death is not captured and there is no linkage with other data sources. Approx. 10.5 Million insurees are included in the database, 7.8 Million of these actively insured in 2024. This corresponds to 7% of the total German population. Data are longitudinally linked over a period of currently ten years.
13	Main references	Andersohn F, Walker J "Characteristics and external validity of the German Health Risk Institute (HRI) Database." Pharmacoepidemiology and drug safety (2016): 26530279
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111207 Website: https://www.ingef.de/en/

The Netherlands: Integrated Primary Care Information (IPCI)

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

#	Section	Description
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content /source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients can be registered under different IDs. However, in the Netherlands, patients typically have one GP and changing practice is uncommon.
9	Quality control (database specific)	Prior to each data release, extensive quality control steps are performed, e.g., comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g. number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical data, e.g. distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.
13	Main references	de Ridder MAJ, de Wilde M,de Ben C,Leyba AR,Mosseveld BMT,Verhamme KMC,van der Lei J,Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182143
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

Norway: Norwegian Linked Health Registry data (NLHR)

#	Section	Description
1	Data source identification and country	NLHR (Norwegian Linked Health Registry data) Norway
2	Data partner information section	University of Oslo Faculty of Mathematics and Natural Science – Department of Pharmacy

#	Section	Description
3	Coverage and timespan	Data coverage: 2008-2023 (primary care data), 2018-2023 (secondary care data and drug dispensation data) Extent: Nation-wide. 5.5M active population. 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.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following registries are included: 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).
5	Data collection process	Registries. 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.
6	General representativeness	The NLHR data covers the full Norwegian population.
7	Data content /source coding	NPR: ICD-10 for diagnosis, ATC and some special codes for drug use, Norwegian codes for clinical procedures (surgery (NCSP), medicine (NCMP) and diagnostic imaging, image-guided intervention, and nuclear medicine (NCRP)). KUHR: ICD-10 and ICPC-2 and ICPC-2B for diagnosis/procedure. NorPD: ATC. SYSVAK and MSIS: national classifications. MBRN: custom classifications by questionnaires (incl. check box variables in Maternity health care card)
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Linkage between the registries was facilitated using project-specific person IDs generated from unique personal identification assigned at birth or immigration for all legal residents in Norway.
9	Quality control (data source specific)	In-house data quality checks of rates of common conditions, drug exposures, and outcomes. We compare obtained rates with official national statistics (e.g., birth statistics, yearly rates of drug dispensing, and diagnosis by age and gender). We also review missing data and outliers and inform registry holders of any unusual patterns.
10	Linkage	The NLHR is, by definition, a linkage of datasets. Helsedata.no is one central portal to apply for 11 national health registries, including all the registries that have been mapped to the OMOP CDM.
11	Vital status	The national death registry is linked.
12	Limitations	(i) Diagnostic codes in the secondary care are limited to the comprehensive list requested from the registries. The list of codes will be revised annually with each data delivery. (ii) ICD-10 codes vary in granularity, ranging from 2-character to full-length codes. The granularity of data will be revised annually with each data delivery. (iii) Drug dispensations to outpatients are included; however, over-the-counter (OTC) medications are not captured in the data. Only a few selected expensive drugs given in hospital, are included.
13	Main references	Trinh et al. Harmonizing Norwegian registries onto OMOP common data model: Mapping challenges and opportunities for pregnancy and COVID-19 research. Int J Med Inform 2024: 191: 105602.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000409 Website: https://www.mn.uio.no/farmasi/english/research/groups/pharma-safe/

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

#	Section	Description
1	Data source identification and country	SIDIAP (The Information System for Research in Primary Care) Catalunya, Spain
2	Data partner information section	IDIAPJGol
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. SIDIAP is a database of primary care electronic health records of the population of Catalonia, North-East Spain. It contains pseudo-anonymised records of more than 8 million people , of which 6.1 million are active as of 2022, representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – General Practitioner, and hospital inpatient care. SIDIAP captured data includes routine visits, socio-demographics, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, sick leaves, and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Catala de la Salut (Catalan Health Institute) is the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content /source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient has a unique id, also upon changing practice or re-registration.
9	Quality control (data source specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g., there are several variables with information concerning the women's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinally and reliability), plausibility (range checks and unusual values), and consistency are inspected through visualisation tools. In addition, before accessing the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences, and prevalence of the data requested for the project, allowing detection of inconsistencies in the data extraction prior to data delivery. External validation processes of the SIDIAP database mainly include assessing the data recorded in SIDIAP through linkage to external gold standard data sources, by analysing free text, or by sending questionnaires to health professionals.

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

Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

#	Section	Description
1	Data source identification and country	HI-SPEED (Health Impact - Swedish Population Evidence Enabling Data-linkage) Sweden
2	Data partner information section	Swedish Medical Products Agency - Gothenburg University (SMPA-GU) Pharmacoepidemiology and Analysis Department (FeA) - School of Public Health and Community Medicine, Institute of Medicine
3	Coverage and timespan	Data collection since: 2015 Extent: Nation-wide. The catchment area includes the whole of Sweden, covering the full population of approximately 11.7 million.
4	Healthcare setting / type of data	Primary care - General Practitioner, and secondary care - specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Primary care (GPs) is available only for the 2 largest regions (~40% of national population) The following data elements are collected: Socio-demographics, dispensed drug prescriptions, cause of death, diagnoses and procedures from secondary (specialist) care and inpatient visits or clinical events, as well as from primary care visits (40%pop only).
5	Data collection process	Registries. The data is acquired from the relevant Swedish national and regional registries, only once all legislative, GDPR and ethical approvals have been granted. Therefore only relevant data is passed on, which will then be entered and processed by the study team. The data are updated several times annually.
6	General representativeness	The coverage includes all patients of all sociodemographic characteristics. Therefore it should mirror the source population to a very good extent.
7	Data content /source coding	Medicines are coded with ATC and NPLID (National Product ID), ICD10-SE is used for diagnoses and the Swedish procedure coding system (KVA) is used for clinical procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients have a uniquely id across datasets.
9	Quality control (data source specific)	The source data are obtained from the relevant Swedish National and Regional Registers. The registers perform some regular quality controls on their data. After receiving the data, we

#	Section	Description
		perform additional checks and cleaning. We also run regular quality checks on the data we manage.
10	Linkage	Data on specialist care is acquired from the National Patient Register, mortality information is provided by the Cause-Of-Death Registry. Drug data is provided by the Patient Drug Register. And similarly for other registers. Data are linked very accurately using the national personal ID number, and pseudonymized before delivery to HI-SPEED. All data are updated 2-4 times per year.
11	Vital status	Data on death and underlying + contributing causes-of-death are extracted from the Cause-of-Death registry (i.e. based on death certificates).
12	Limitations	General limitations for the data type applicable. This is a research project where all studies require ethics approval. Data collection since: 2015 for most data, except prescribed drug register (from 2018), and some COVID-related data (tests, vaccination) from 2020 Primary care is only available for a subset.
13	Main references	Nyberg F, Franzén S,Lindh M,Vanfleteren L,Hammar N,Wettermark B,Sundström J,Santosa A,Björck S,Gisslén M "Swedish Covid-19 Investigation for Future Insights – A Population Epidemiology Approach Using Register Linkage (SCIFI-PEARL)." Clinical epidemiology (2021): 34354377
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/node/4463/ Website: https://www.gu.se/en/research/scifi-pearl

The UK: Clinical Practice Research Datalink GOLD (CPRD GOLD)

#	Section	Description
1	Data source identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) The United Kingdom
2	Data partner information section	University of Oxford NDORMS
3	Coverage and timespan	Data collection since: 1987 Extent: Nation-wide. CPRD GOLD consists of patients in contributing practices using Vision software. Historically this covered the whole of the UK, but the number of contributing practices in the England is dropping. In January 2025 only 3 practices from England were a part of CPRD GOLD, while historical patient data were from the whole of the UK, and will continue to be so. In the future, no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records. Data are entered by clinicians into the EHR. Data is processed by CPRD that provides data releases for research.
6	General representativeness	In the last 10 years, the CPRD GOLD regional distribution of currently contributing general practitioner (GP) practices has significantly shifted, resulting in many new practices joining from Scotland, Wales, and Northern Ireland, and fewer participating from England. These changes have affected the CPRD GOLD population size, regional coverage, and eligibility for data linkages. CPRD GOLD January 2024 contains >21.3 million historical and current patients (12.9 in England, 3.1 in Wales, 4.7 in Scotland, 0.7 in Northern Ireland). Of these, nearly 3 million are currently

#	Section	Description
		registered in a GP practice and represent ~4.3% of the estimated current UK population (0.1% in England, 32.3% in Wales, 28.6% in Scotland, 16.2% in Northern Ireland). Patients currently registered in CPRD GOLD January 2024 are broadly representative of the UK population with respect to age and sex. Reference: https://doi.org/10.1093/ije/dyaf077
7	Data content /source coding	Gemscript, Read, dm+d
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In GOLD, a patient can be registered under different ID numbers upon changing practice or re-registration. Researchers are not able to identify these patients, as the data are anonymised. However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (data source specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (UTS). Each practice is associated to an UTS date set when the data quality standards become satisfactory, and CPRD recommend using only longitudinal data starting from this UTS date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards, and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice UTS date. CPRD also checks data quality standards at the patient level, and associates each patient with a flag, reporting if its data are acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at the Practice Level.
11	Vital status	The date of death in CPRD GOLD has been validated against the Population registry (ONS) mortality data. Reference: https://doi.org/10.1002/pds.4747
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL,Dedman D,Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://www.cprd.com/data/primary-care-data/cprd-gold

ANNEX II. Fitness for use assessment

Danish Data Health Registries (DK-DHR)

DK-DHR was included in this study because it is a secondary care and hospital data source that provided relevant information on AESIs in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for AESIs (All objectives) in DK-DHR was 1.41M.

Moreover, data availability and follow-up in DK-DHR was sufficient, as data availability starts in 1995, and the date of most recent data extraction was 18th January 2025, which aligned with the study period. The median follow-up of the first observation period in DK-DHR was 7,920 days.

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

Lastly, DK-DHR had blanket approval, which made the execution of this study feasible within the study timelines.

Finnish Care Register for Health Care (FinOMOP-THL)

FinOMOP-THL was included in this study because it is a primary care and hospital data source that provided relevant information on AESIs in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for AESIs (All objectives) in FinOMOP-THL was 968.3k.

Moreover, data availability and follow-up in FinOMOP-THL was sufficient, as data availability starts in 2011, and the date of most recent data extraction was 1st October 2024, which aligned with the study period. The median follow-up of the first observation period in FinOMOP-THL was 5,020 days.

There were study specific limitations present in FinOMOP-THL, namely: Diagnosis from private healthcare sector may only have observation in the last year of follow-up.

Lastly, IRB approval for FinOMOP-THL was estimated to take a week from submission, which made the execution of this study feasible within the study timelines.

InGef Research Database (InGef RDB)

InGef RDB was included in this study because it is a claims data source that provided relevant information AESIs in the general population from in-patient and out-patient hospital settings.

Based on a preliminary feasibility assessment, the expected number of person-counts for AESIs (All objectives) in InGef RDB was 1.22M.

Moreover, data availability and follow-up in InGef RDB was sufficient, as data availability starts in 2016, and the date of most recent data extraction was 2nd Jan 2026, which aligned with the study period. The median follow-up of the first observation period in InGef RDB was 3,560 days.

There were study specific limitations present in InGef RDB, namely: absence of data on outcome of interest prior to 2016 (All objectives), due to data availability the study period for InGef RDB was 1st January 2016 to study period end date.

Lastly, IRB approval for InGef RDB was estimated to take 2–4 weeks, which made the execution of this study feasible within the study timelines.

Integrated Primary Care Information (IPCI)

IPCI was included in this study because it is a primary care data source that provided relevant information on AESIs in the general population, since it has nationwide representativeness.

Based on a preliminary feasibility assessment, the expected number of person-counts for AESIs (All objectives) in IPCI was 236.8k.

Moreover, data availability and follow-up in IPCI was sufficient, as data availability starts in 2006, and the date of most recent data extraction was 1st December 2024, which aligned with the study period. The median follow-up of the first observation period in IPCI was 1,730 days.

There were some study specific limitations present in IPCI, namely: the potential under capturing of diagnosis of AESIs cases, if the diagnosis is made primarily in hospital.

Lastly, IPCI had blanket approval, which made the execution of this study feasible within the study timelines.

Norwegian Linked Health Registry (NLHR)

NLHR was included in this study because it is a registry data source that provided relevant information on AESIs in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for AESIs (All objectives) in NLHR was 720.5k.

Moreover, data availability and follow-up in NLHR was sufficient, as data availability starts in 2008, and the date of most recent data extraction was 29th October 2024, which aligned with the study period. The median follow-up of the first observation period in NLHR was 5,820 days.

There were no study specific limitations present in NLHR.

Lastly, NLHR had blanket approval, which made the execution of this study feasible within the study timelines.

Health Impact – Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

HI-SPEED was included in this study because it is a registry data source that provided relevant information on AESIs in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for AESIs (All objectives) in HI-SPEED was 938.7k.

Moreover, data availability and follow-up in HI-SPEED was sufficient, as data availability starts in 2015, and the date of most recent data extraction was 31st August 2025, which aligned with the study period. The median follow-up of the first observation period in HI-SPEED was 3,530 days.

There were some study specific limitations present in HI-SPEED, namely: absence of data on outcome of interest prior to 2015 (All objectives), due to data availability the study period for HI-SPEED was 1st January 2015 (1st January 2016 with 365 days of prior data availability prior to cohort entry data requirement) to study period end date.

Lastly, ethics approval for HI-SPEED was estimated to take 30 days from the ethics submission of protocol, which made the execution of this study feasible within the study timelines.

The Information System for Research in Primary Care (SIDIAP)

SIDIAP was included in this study because it is a primary care and hospital data source that provided relevant information on AESIs in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for AESIs (All objectives) in SIDIAP was 794k.

Moreover, data availability and follow-up in SIDIAP was sufficient, as data availability starts in 2006, and the date of most recent data extraction was 31st December 2024, which aligned with the study period. The median follow-up of the first observation period in SIDIAP was 5,670 days.

There were some study specific limitations present in SIDIAP, namely: Although it includes hospital data, it only includes information recorded in the discharge report after hospital admission. Conditions managed entirely in the hospital emergency room are not captured.

Lastly, IRB approval for SIDIAP was estimated to take 2 months, which made the execution of this study feasible within the study timelines.

Clinical Practice Research Datalink GOLD (CPRD GOLD)

CPRD GOLD was included in this study because it is a primary care data source that provided relevant information on AESIs in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for AESIs (All objectives) in CPRD GOLD was 811.2k.

Moreover, data availability and follow-up in CPRD GOLD was sufficient, as data availability starts in 1987, and the date of most recent data extraction was 1st July 2025, which aligned with the study period. The median follow-up of the first observation period in CPRD GOLD was 2,150 days.

There were some study specific limitations present in CPRD GOLD, namely: the potential under capturing of diagnosis of AESIs cases, if the diagnosis is made primarily in hospital.

Lastly, CPRD GOLD had blanket approval, which made the execution of this study feasible within the study timelines.

ANNEX III. Operational and reporting considerations

DATA MANAGEMENT

Data management

All data sources have previously mapped their data to the OMOP common data model. This enabled the use of standardised analytics and using DARWIN EU® tools across the network since the structure of the data and the terminology system was 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 data source containing individual data and then returned the results (csv files) which only contained aggregated data. The results from each of the contributing data sites were then combined in tables and figures for the study report.

Data storage and protection

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

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

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

QUALITY CONTROL

Data source quality control

When defining cohorts for indications, a systematic search of possible codes for inclusion were identified using the *CodelistGenerator* R package (https://github.com/darwin-eu/CodelistGenerator). This package allows the user to define a search strategy and will use this to query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, the *CohortDiagnostics* (https://github.com/OHDSI/CohortDiagnostics) R packages was ran, if needed, to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error.

The study code was based on the DARWIN EU® R packages *IncidencePrevalence*[11] to estimate Incidence and Prevalence. This package will include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package was made publicly available via GitHub.

ANNEX IV. List of stand-alone documents

Table S1. Codelist of Guillain-Barré syndrome definitions.

Concept ID	Concept name	Domain	Vocabulary
4164770	Guillain-Barré syndrome	Condition	SNOMED
37396786	Acute inflammatory demyelinating polyradiculoneuropathy	Condition	SNOMED
4070552	Fisher's syndrome	Condition	SNOMED
4141686	Bickerstaff's brainstem encephalitis	Condition	SNOMED
43530713	Acute inflammatory demyelinating polyneuropathy	Condition	SNOMED

Table S2. Codelist of immune thrombocytopaenia definitions.

Concept ID	Concept name	Domain	Vocabulary
4027374	Alloimmune platelet transfusion refractoriness	Condition	SNOMED
4133984	Alloimmune thrombocytopenia	Condition	SNOMED
318397	Chronic idiopathic thrombocytopenic purpura	Condition	SNOMED
4000065	Drug-induced immune thrombocytopenia	Condition	SNOMED
4119134	Thrombocytopenic purpura	Condition	SNOMED
4103532	Immune thrombocytopenia	Condition	SNOMED
4028065	Primary ITP (immune thrombocytopenia)	Condition	SNOMED
4133983	Secondary autoimmune thrombocytopenia	Condition	SNOMED
4102469	Acute idiopathic thrombocytopenic purpura	Condition	SNOMED
4137430	Idiopathic thrombocytopenic purpura	Condition	SNOMED
441264	Primary thrombocytopenia	Condition	SNOMED
436956	Evans syndrome	Condition	SNOMED

Table S3. Codelist of myocarditis definitions.

Concept ID	Concept name	Domain	Vocabulary
312653	Acute myocarditis	Condition	SNOMED
4147350	Acute myoendocarditis	Condition	SNOMED
4006483	Acute myopericarditis	Condition	SNOMED
317307	Idiopathic myocarditis	Condition	SNOMED
43020652	Inflammatory cardiomyopathy	Condition	SNOMED
4242712	Interstitial myocarditis	Condition	SNOMED
314383	Myocarditis	Condition	SNOMED
4219260	Myocarditis due to drug	Condition	SNOMED
4336533	Myocarditis due to hypersensitivity state	Condition	SNOMED
37109914	Myopericarditis	Condition	SNOMED
4143969	Isolated (Fiedler's) myocarditis	Condition	SNOMED
4071283	Acute benign pericarditis	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4072086	Acute bloody pericarditis	Condition	SNOMED
4070323	Acute effusive pericarditis	Condition	SNOMED
315293	Acute idiopathic pericarditis	Condition	SNOMED
4304392	Acute mediastinopericarditis	Condition	SNOMED
320116	Acute pericarditis	Condition	SNOMED
4111399	Acute pericarditis secondary to uremia	Condition	SNOMED
4204985	Acute periendocarditis	Condition	SNOMED
4069712	Acute pleuropericarditis	Condition	SNOMED
4111080	Acute pneumococcal pericarditis	Condition	SNOMED
4178425	Acute pneumopericarditis	Condition	SNOMED
321586	Acute rheumatic pericarditis	Condition	SNOMED
4108799	Acute tuberculous pericarditis	Condition	SNOMED
44782774	Chest pain due to pericarditis	Condition	SNOMED
315469	Coxsackie pericarditis	Condition	SNOMED
4031469	Drug-induced pericarditis	Condition	SNOMED
312543	Meningococcal pericarditis	Condition	SNOMED
4218568	Non-infectious pericarditis	Condition	SNOMED
4196638	Pericarditis secondary to uremia	Condition	SNOMED
4245386	Pericarditis sicca	Condition	SNOMED
4121477	Post-infarction pericarditis	Condition	SNOMED
4289908	Viral pericarditis	Condition	SNOMED
312928	adhesive pericarditis	Condition	SNOMED
4149913	Systemic lupus erythematosus with pericarditis	Condition	SNOMED
4217075	Infectious pericarditis	Condition	SNOMED
37166030	Post-cardiac injury syndrome	Condition	SNOMED
4121477	Post-infarction pericarditis	Condition	SNOMED

Table S4. Codelist of ischaemic stroke definitions.

Concept ID	Concept name	Domain	Vocabulary
4048784	Anterior cerebral circulation hemorrhagic infarction	Condition	SNOMED
4045735	Anterior cerebral circulation infarction	Condition	SNOMED
4031045	Anterior choroidal artery syndrome	Condition	SNOMED
761110	Bilateral cerebral infarction due to precererbral arterial occlusion	Condition	SNOMED
4319331	Cerebellar infarction	Condition	SNOMED
372924	Cerebral artery occlusion	Condition	SNOMED
4110189	Cerebral infarct due to thrombosis of precerebral arteries	Condition	SNOMED
443454	Cerebral infarction	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
762951	Cerebral infarction due to anterior cerebral artery occlusion	Condition	SNOMED
765515	Cerebral infarction due to basilar artery stenosis	Condition	SNOMED
43530683	Cerebral infarction due to carotid artery occlusion	Condition	SNOMED
762933	Cerebral infarction due to cerebral artery occlusion	Condition	SNOMED
762937	Cerebral infarction due to cerebral venous thrombosis	Condition	SNOMED
4108356	Cerebral infarction due to embolism of cerebral arteries	Condition	SNOMED
45772786	Cerebral infarction due to embolism of middle cerebral artery	Condition	SNOMED
4110190	Cerebral infarction due to embolism of precerebral arteries	Condition	SNOMED
762935	Cerebral infarction due to internal carotid artery occlusion	Condition	SNOMED
763015	Cerebral infarction due to middle cerebral artery occlusion	Condition	SNOMED
46273649	Cerebral infarction due to occlusion of basilar artery	Condition	SNOMED
35610084	Cerebral infarction due to occlusion of cerebral artery	Condition	SNOMED
46270031	Cerebral infarction due to occlusion of precerebral artery	Condition	SNOMED
762934	Cerebral infarction due to posterior cerebral artery occlusion	Condition	SNOMED
43531607	Cerebral infarction due to stenosis of carotid artery	Condition	SNOMED
35610085	Cerebral infarction due to stenosis of cerebral artery	Condition	SNOMED
46270381	Cerebral infarction due to stenosis of precerebral artery	Condition	SNOMED
4110192	Cerebral infarction due to thrombosis of cerebral arteries	Condition	SNOMED
45767658	Cerebral infarction due to thrombosis of middle cerebral artery	Condition	SNOMED
44782773	Cerebral infarction due to vertebral artery occlusion	Condition	SNOMED
46270380	Cerebral infarction due to vertebral artery stenosis	Condition	SNOMED
37110678	Cerebral ischemic stroke due to occlusion of extracranial large artery	Condition	SNOMED
37110679	Cerebral ischemic stroke due to stenosis of extracranial large artery	Condition	SNOMED
4045734	CVA – cerebrovascular accident due to cerebral artery occlusion	Condition	SNOMED
4046362	Hemorrhagic cerebral infarction	Condition	SNOMED
4131383	Infarction of basal ganglia	Condition	SNOMED
4046237	Infarction of optic radiation	Condition	SNOMED
4119140	Infarction of visual cortex	Condition	SNOMED
4043731	Infarction – precerebral	Condition	SNOMED
4046360	Lacunar infarction	Condition	SNOMED
4141405	Left sided cerebral infarction	Condition	SNOMED
37116473	Multifocal cerebral infarction due to and following procedure on cardiovascular system	Condition	SNOMED
4077086	Occipital cerebral infarction	Condition	SNOMED
4046359	Partial anterior cerebral circulation infarction	Condition	SNOMED
4319146	Pituitary infarction	Condition	SNOMED
4043732	Posterior cerebral circulation hemorrhagic infarction	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4045737	Pure motor lacunar infarction	Condition	SNOMED
4146185	Right sided cerebral infarction	Condition	SNOMED
36717605	Silent cerebral infarct	Condition	SNOMED
4142739	Thalamic infarction	Condition	SNOMED
4046358	Total anterior cerebral circulation infarction	Condition	SNOMED
4045738	Pure sensory lacunar infarction	Condition	SNOMED
40479572	Infarct of cerebrum due to iatrogenic cerebrovascular accident	Condition	SNOMED
44782781	Hemiplegia and/or hemiparesis following stroke	Condition	SNOMED
40481354	Speech and language deficit as late effect of cerebrovascular accident	Condition	SNOMED
4112023	Occlusion and stenosis of middle cerebral artery	Condition	SNOMED
600670	Occlusion of anterior cerebral artery	Condition	SNOMED
4111717	Occlusion of posterior cerebral artery	Condition	SNOMED
441874	Cerebral thrombosis	Condition	SNOMED
375557	Cerebral embolism	Condition	SNOMED

Table S5. Codelist of haemorrhagic stroke definitions.

Concept ID	Concept name	Domain	Vocabulary
35609033	Haemorrhagic stroke	Condition	SNOMED
42535421	Spontaneous hemorrhage of subarachnoid space from intracranial artery	Condition	SNOMED
45766118	Non-aneurysmal subarachnoid intracranial hemorrhage	Condition	SNOMED
42538062	Spontaneous intracranial hemorrhage	Condition	SNOMED
42873042	Subpial intracranial hemorrhage	Condition	SNOMED
4172545	Angiogram-negative subarachnoid hemorrhage	Condition	SNOMED
432923	Subarachnoid hemorrhage	Condition	SNOMED
4046365	Subarachnoid hemorrhage due to ruptured aneurysm	Condition	SNOMED
4111708	Subarachnoid hemorrhage from vertebral artery	Condition	SNOMED
4148906	Spontaneous subarachnoid hemorrhage	Condition	SNOMED
42535423	Spontaneous hemorrhage of deep cerebral hemisphere	Condition	SNOMED
42535424	Spontaneous hemorrhage of cortical intracerebral hemisphere	Condition	SNOMED
42535425	Spontaneous hemorrhage of cerebral hemisphere	Condition	SNOMED
42539269	Spontaneous hemorrhage of brain stem	Condition	SNOMED
43530674	Spontaneous cerebellar hemorrhage	Condition	SNOMED
43530727	Spontaneous cerebral hemorrhage	Condition	SNOMED
44782730	Nontraumatic intraparenchymal cerebral hemorrhage	Condition	SNOMED
45766120	Convexal subarachnoid hemorrhage	Condition	SNOMED
46270111	Nontraumatic subarachnoid hemorrhage with brain compression	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
37309659	Spontaneous hemorrhage of subarachnoid space from anterior communicating artery	Condition	SNOMED
46273491	Spontaneous cerebral hemorrhage with compression of brain	Condition	SNOMED
42535420	Spontaneous hemorrhage of subarachnoid space from basilar artery	Condition	SNOMED
4077200	Subarachnoid hemorrhage from posterior cerebral artery aneurysm	Condition	SNOMED
42535422	Spontaneous hemorrhage of subarachnoid space from left posterior communicating artery	Condition	SNOMED
4077201	Subarachnoid hemorrhage from basilar artery aneurysm	Condition	SNOMED
4077827	Subarachnoid hemorrhage from multiple aneurysms	Condition	SNOMED
4077828	Subarachnoid hemorrhage from anterior cerebral artery aneurysm	Condition	SNOMED
42539183	Spontaneous hemorrhage of subarachnoid space from right posterior communicating artery	Condition	SNOMED
4077958	Subarachnoid hemorrhage from anterior communicating artery aneurysm	Condition	SNOMED
4077959	Subarachnoid hemorrhage from posterior communicating artery aneurysm	Condition	SNOMED
4078446	Subarachnoid hemorrhage from middle cerebral artery aneurysm	Condition	SNOMED
762095	Spontaneous hemorrhage of subarachnoid space from left middle cerebral artery	Condition	SNOMED
4078447	Subarachnoid hemorrhage from posterior inferior cerebellar artery aneurysm	Condition	SNOMED
762096	Non-traumatic hemorrhage of subarachnoid space from right middle cerebral artery	Condition	SNOMED
4046364	Subarachnoid hemorrhage due to ruptured arteriovenous malformation	Condition	SNOMED
4078448	Subarachnoid hemorrhage from carotid artery aneurysm	Condition	SNOMED
4108952	Subarachnoid hemorrhage from carotid siphon and bifurcation	Condition	SNOMED
4173481	Cerebromeningeal hemorrhage	Condition	SNOMED
45766119	Perimesencephalic subarachnoid hemorrhage	Condition	SNOMED
45766830	Non-aneurysmal perimesencephalic subarachnoid hemorrhage	Condition	SNOMED
4144154	Non-traumatic intracerebral ventricular hemorrhage	Condition	SNOMED
45773167	Subarachnoid hemorrhage from vertebral artery aneurysm	Condition	SNOMED
376713	Cerebral hemorrhage	Condition	SNOMED
443752	Ventricular hemorrhage	Condition	SNOMED
4045743	Massive supratentorial cerebral hemorrhage	Condition	SNOMED
4045744	Lobar cerebral hemorrhage	Condition	SNOMED
4045745	Thalamic hemorrhage	Condition	SNOMED
4049659	Subcortical hemorrhage	Condition	SNOMED
4080892	Intracerebellar and posterior fossa hemorrhage	Condition	SNOMED
4110185	Intracerebral hemorrhage, intraventricular	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4110186	Intracerebral hemorrhage, multiple localized	Condition	SNOMED
4112018	Basal ganglia hemorrhage	Condition	SNOMED
4112019	External capsule hemorrhage	Condition	SNOMED
4129535	Pituitary hemorrhage	Condition	SNOMED
4176892	Cortical hemorrhage	Condition	SNOMED
4178726	Internal capsule hemorrhage	Condition	SNOMED
37109909	Silent micro-hemorrhage of brain	Condition	SNOMED
4199890	Hematoma of brain	Condition	SNOMED
4201094	Cerebellar hematoma	Condition	SNOMED
4218781	Cerebral hemisphere hemorrhage	Condition	SNOMED
37116267	Hemorrhage of medulla oblongata	Condition	SNOMED
4298750	Periventricular hemorrhagic venous infarct	Condition	SNOMED
40492969	Intraparenchymal hemorrhage of brain	Condition	SNOMED
4299377	Intrapontine hemorrhage	Condition	SNOMED
4319328	Brain stem hemorrhage	Condition	SNOMED
4326561	Cerebellar hemorrhage	Condition	SNOMED
4328027	Intraparenchymal hematoma of brain	Condition	SNOMED
42872434	Intracranial hematoma	Condition	SNOMED
42873044	Hemorrhage in globus pallidus	Condition	SNOMED
42873045	Hemorrhage in caudate nucleus	Condition	SNOMED
42873046	Hemorrhage in putamen	Condition	SNOMED
45766068	Deep hemispheric cerebral hemorrhage	Condition	SNOMED
762351	Ruptured acquired aneurysm of cerebral artery	Condition	SNOMED
764707	Ruptured aneurysm of intracranial artery	Condition	SNOMED

Table S6. Codelist of Encephalitis definitions.

Concept ID	Concept name	Domain	Vocabulary
36676401	Post-transplant acute limbic encephalitis	Condition	SNOMED
36676495	Limbic encephalitis with dipeptidyl-peptidase 6 antibodies	Condition	SNOMED
36715969	Encephalitis caused by Alphavirus	Condition	SNOMED
36715978	Encephalitis caused by Polyoma virus	Condition	SNOMED
36715986	Encephalitis caused by Rubulavirus	Condition	SNOMED
36715988	Encephalitis caused by Hendra virus	Condition	SNOMED
36716007	Encephalitis caused by Trypanosoma brucei	Condition	SNOMED
36716010	Encephalitis caused by Schistosoma	Condition	SNOMED
36716011	Encephalitis caused by Schistosoma haematobium	Condition	SNOMED
36716012	Encephalitis caused by Schistosoma mansoni	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
36716013	Encephalitis caused by Schistosoma japonicum	Condition	SNOMED
36716014	Encephalitis caused by Echinococcus granulosus	Condition	SNOMED
36716016	Encephalitis caused by Taenia solium	Condition	SNOMED
36716017	Encephalitis caused by Coenurus cerebralis	Condition	SNOMED
36717188	Encephalitis caused by Nipah virus	Condition	SNOMED
37016847	Encephalitis caused by tick-borne encephalitis virus	Condition	SNOMED
37017071	Subacute adenoviral encephalitis co-occurrent with human immunodeficiency virus infection	Condition	SNOMED
37116467	Dementia due to herpes encephalitis	Condition	SNOMED
37311983	Henipavirus encephalitis	Condition	SNOMED
37399546	Limbic encephalitis with N-methyl-D-aspartate receptor antibodies	Condition	SNOMED
40479557	Encephalitis due to human herpesvirus 6 infection	Condition	SNOMED
40481971	Paraneoplastic limbic encephalitis	Condition	SNOMED
40483321	Bacterial meningoencephalitis	Condition	SNOMED
44784560	Dementia associated with viral encephalitis	Condition	SNOMED
44806551	Acute encephalitis	Condition	SNOMED
45757204	Encephalitis due to Actinomyces	Condition	SNOMED
45757219	Meningoencephalitis due to Blastomyces dermatitidis	Condition	SNOMED
46269919	Meningoencephalitis due to Acanthamoeba	Condition	SNOMED
46269920	Meningoencephalitis due to Chagas disease	Condition	SNOMED
46270062	Meningoencephalitis due to free-living ameba	Condition	SNOMED
372547	Viral encephalitis	Condition	SNOMED
372615	Post-infectious encephalitis	Condition	SNOMED
373408	Subacute sclerosing panencephalitis	Condition	SNOMED
373692	West Nile encephalitis	Condition	SNOMED
374020	Infective encephalitis	Condition	SNOMED
375475	Meningoencephalitis due to acquired toxoplasmosis	Condition	SNOMED
376026	Encephalitis due to Herpesviridae	Condition	SNOMED
376905	Encephalitis due to Far Eastern tick-borne encephalitis virus	Condition	SNOMED
377487	Post measles encephalitis	Condition	SNOMED
378065	Mumps encephalitis	Condition	SNOMED
378136	Toxic encephalitis	Condition	SNOMED
378143	Encephalitis	Condition	SNOMED
378349	Syphilitic encephalitis	Condition	SNOMED
379792	Post-immunization encephalitis	Condition	SNOMED
380322	Postvaricella encephalitis	Condition	SNOMED
380632	Meningococcal encephalitis	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
380941	Herpetic meningoencephalitis	Condition	SNOMED
380942	Western equine encephalitis	Condition	SNOMED
381789	Japanese encephalitis virus disease	Condition	SNOMED
381794	Congenital syphilitic encephalitis	Condition	SNOMED
439725	California encephalitis virus infection	Condition	SNOMED
440023	St. Louis encephalitis virus infection	Condition	SNOMED
440332	Eastern equine encephalitis virus infection	Condition	SNOMED
442611	Murray Valley encephalitis	Condition	SNOMED
443543	Encephalitis due to rickettsia	Condition	SNOMED
443579	St. Louis encephalitis virus non-neuroinvasive disease	Condition	SNOMED
443881	Encephalitis due to European subtype of tick-borne encephalitis virus	Condition	SNOMED
762645	Meningoencephalitis caused by virus	Condition	SNOMED
438069	Venezuelan equine encephalitis virus infection	Condition	SNOMED
764228	Autoimmune encephalitis caused by N-methyl D-aspartate receptor antibody	Condition	SNOMED
4002949	Progressive rubella panencephalitis	Condition	SNOMED
4009332	Acute necrotizing encephalitis	Condition	SNOMED
4009925	Subacute adenoviral encephalitis	Condition	SNOMED
4027838	Rubella meningoencephalitis	Condition	SNOMED
4041669	Enteroviral encephalitis	Condition	SNOMED
4041670	Infectious mononucleosis encephalitis	Condition	SNOMED
4041671	Brainstem encephalitis	Condition	SNOMED
4041673	Limbic encephalitis	Condition	SNOMED
4042197	Rhabdovirus encephalitis	Condition	SNOMED
4042200	Primary amebic encephalitis	Condition	SNOMED
4042201	Granulomatous amebic encephalitis	Condition	SNOMED
4042202	Post-influenza encephalitis	Condition	SNOMED
4044936	Tensaw encephalitis	Condition	SNOMED
4044937	Kumlinge virus encephalitis	Condition	SNOMED
4045976	Herpes zoster encephalitis	Condition	SNOMED
4045977	Bacterial encephalitis	Condition	SNOMED
4045978	Late syphilitic encephalitis	Condition	SNOMED
4045979	Amebic encephalitis	Condition	SNOMED
4047470	Chronic echovirus meningoencephalitis	Condition	SNOMED
4047619	Acute viral encephalitis	Condition	SNOMED
4047620	Rocio virus encephalitis	Condition	SNOMED
4047622	Herpes simiae encephalitis	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4047623	Chronic viral encephalitis	Condition	SNOMED
4047625	Fungal encephalitis	Condition	SNOMED
4047626	Focal encephalitis	Condition	SNOMED
4077718	Listeria meningoencephalitis	Condition	SNOMED
4088248	Epidemic encephalitis	Condition	SNOMED
4088249	Encephalitis lethargica	Condition	SNOMED
4100107	Rubella encephalitis	Condition	SNOMED
4100108	Post tetanus vaccination encephalitis	Condition	SNOMED
4100109	Post smallpox vaccination encephalitis	Condition	SNOMED
4100110	Post typhus vaccination encephalitis	Condition	SNOMED
4100111	Post rubella vaccination encephalitis	Condition	SNOMED
4100112	Post hepatitis A vaccination encephalitis	Condition	SNOMED
4102195	Arbovirus encephalitis	Condition	SNOMED
4102198	Post BCG vaccination encephalitis	Condition	SNOMED
4102199	Post measles vaccination encephalitis	Condition	SNOMED
4103103	Polioencephalitis	Condition	SNOMED
4103105	Toxoplasma encephalitis	Condition	SNOMED
4103107	Post typhoid vaccination encephalitis	Condition	SNOMED
4103108	Post paratyphoid vaccination encephalitis	Condition	SNOMED
4103109	Post cholera vaccination encephalitis	Condition	SNOMED
4103110	Post plague vaccination encephalitis	Condition	SNOMED
4103111	Post pertussis vaccination encephalitis	Condition	SNOMED
4103112	Post rabies vaccination encephalitis	Condition	SNOMED
4103113	Post yellow fever vaccination encephalitis	Condition	SNOMED
4103114	Post polio vaccination encephalitis	Condition	SNOMED
4103116	Post influenza vaccination encephalitis	Condition	SNOMED
4104690	Post hepatitis B vaccination encephalitis	Condition	SNOMED
4105200	Post diphtheria vaccination encephalitis	Condition	SNOMED
4105201	Post mixed vaccination encephalitis	Condition	SNOMED
4141686	Bickerstaff's brainstem encephalitis	Condition	SNOMED
4146943	Encephalitis due to influenza	Condition	SNOMED
4147498	Encephalitis, myelitis and encephalomyelitis	Condition	SNOMED
4147586	Acute adenoviral encephalitis	Condition	SNOMED
4150516	Tuberculous meningoencephalitis	Condition	SNOMED
4161950	Human immunodeficiency virus encephalitis	Condition	SNOMED
4166858	Measles inclusion body encephalitis	Condition	SNOMED
4169939	Neuroinvasive Cache Valley encephalitis virus infection	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4174766	Lymphocytic meningoencephalitis	Condition	SNOMED
4176017	Primary amebic encephalitis due to Naegleria fowleri	Condition	SNOMED
4180433	Meningoencephalitis due to Naegleria	Condition	SNOMED
4210015	Eosinophilic meningoencephalitis	Condition	SNOMED
4210628	Adenoviral encephalitis	Condition	SNOMED
4213372	California encephalitis virus infection neuroinvasive disease	Condition	SNOMED
4213917	Rio Bravo viral encephalitis	Condition	SNOMED
4214943	Neuroinvasive St. Louis encephalitis virus infection	Condition	SNOMED
4215521	Powassan encephalitis virus infection	Condition	SNOMED
4217555	Siberian tick-borne encephalitis	Condition	SNOMED
4219689	Neuroinvasive Powassan encephalitis virus infection	Condition	SNOMED
4220632	Mumps meningoencephalitis	Condition	SNOMED
4221064	Eastern equine encephalitis virus neuroinvasive disease	Condition	SNOMED
4221192	Eastern equine encephalitis virus non-neuroinvasive disease	Condition	SNOMED
4221266	Acute adenoviral meningoencephalitis	Condition	SNOMED
4221661	Encephalitis due to Langat virus	Condition	SNOMED
4222250	Encephalitis associated with AIDS	Condition	SNOMED
4222717	Subacute adenoviral encephalitis associated with AIDS	Condition	SNOMED
4227852	Jamestown Canyon virus encephalitis	Condition	SNOMED
4227853	Snowshoe hare virus encephalitis	Condition	SNOMED
4230601	Keystone virus encephalitis	Condition	SNOMED
4233538	Van Bogaert's sclerosing leukoencephalitis	Condition	SNOMED
4236144	Nairoviral encephalitis	Condition	SNOMED
4238921	Tuberculous encephalitis	Condition	SNOMED
4249132	Primary encephalitis	Condition	SNOMED
4252714	Q fever encephalitis	Condition	SNOMED
4265323	La Crosse encephalitis	Condition	SNOMED
4289603	Negishi encephalitis	Condition	SNOMED
4296460	Herpetic acute necrotizing encephalitis	Condition	SNOMED
4302794	Encephalitis due to protozoa	Condition	SNOMED
4302798	Eosinophilic meningoencephalitis due to Angiostrongylus cantonensis	Condition	SNOMED
4306458	Cytomegalovirus encephalitis	Condition	SNOMED
4309568	Non-neuroinvasive Western equine encephalitis virus infection	Condition	SNOMED
4310991	Neuroinvasive Western equine encephalitis virus infection	Condition	SNOMED
4312671	Paraneoplastic encephalitis	Condition	SNOMED
4318558	Autoimmune encephalitis	Condition	SNOMED
4318863	Systemic lupus erythematosus encephalitis	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4322568	Encephalitis due to human herpes simplex virus	Condition	SNOMED
4322814	Meningoencephalitis	Condition	SNOMED
4324130	Trypanosomiasis with encephalitis	Condition	SNOMED
4324289	Allergic encephalitis	Condition	SNOMED
4324541	Ilheus virus encephalitis	Condition	SNOMED
4324751	Primary viral encephalitis	Condition	SNOMED
4327502	Toxic encephalitis due to thallium	Condition	SNOMED
375185	Encephalomyelitis due to rubella	Condition	SNOMED
35607961	Limbic encephalitis with leucine-rich glioma-inactivated 1 antibodies	Condition	SNOMED
35622353	Limbic encephalitis with contactin-associated protein-like 2 antibodies	Condition	SNOMED
35622800	Mycoplasma pneumoniae encephalitis	Condition	SNOMED
35622805	Non-herpetic acute limbic encephalitis	Condition	SNOMED
4216883	Powassan virus non-neuroinvasive disease	Condition	SNOMED
4035637	Simian B encephalomyelitis	Condition	SNOMED
4120303	Progressive congenital rubella encephalomyelitis	Condition	SNOMED
4229086	Post-encephalitic syndrome	Condition	SNOMED
4249574	Acute hemorrhagic leukoencephalitis	Condition	SNOMED
4320933	Enteroviral encephalomyelitis	Condition	SNOMED
36674282	Steroid-responsive encephalopathy associated with autoimmune thyroiditis	Condition	SNOMED
4215383	Neuroinvasive Venezuelan equine encephalitis virus infection	Condition	SNOMED
379798	Post-infectious encephalomyelitis	Condition	SNOMED
4220454	Venezuelan equine encephalitis virus non-neuroinvasive disease	Condition	SNOMED
43530701	Acute disseminated encephalomyelitis following infectious disease	Condition	SNOMED
36715541	Infection causing encephalomyelitis	Condition	SNOMED
40480147	Toxic encephalomyelitis	Condition	SNOMED
373189	Encephalomyelitis	Condition	SNOMED
374021	Acute disseminated encephalomyelitis	Condition	SNOMED
443802	Postvaccinal encephalomyelitis	Condition	SNOMED
761989	Recurrent acute disseminated encephalomyelitis	Condition	SNOMED
763040	Rasmussen syndrome, refractory	Condition	SNOMED
765827	Multiphasic acute disseminated encephalomyelitis	Condition	SNOMED
4176250	Allergic encephalomyelitis	Condition	SNOMED
4206037	Pyogranulomatous meningoencephalomyelitis	Condition	SNOMED
4220014	Encephalomyelitis caused by lymphocytic choriomeningitis virus	Condition	SNOMED
4228612	Encephalomyelitis associated with AIDS	Condition	SNOMED
4310021	Paraneoplastic encephalomyelitis	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4104546	Meningoencephalomyelitis	Condition	SNOMED
4041672	Rasmussen syndrome	Condition	SNOMED
4302073	California serogroup virus neuroinvasive disease	Condition	SNOMED
35607947	Acute necrotizing encephalopathy of childhood	Condition	SNOMED
4252885	Influenza with encephalopathy	Condition	SNOMED
46285143	Bacterial meningoencephalomyelitis	Condition	SNOMED
4242575	Granulomatous meningoencephalomyelitis	Condition	SNOMED
4103115	Post mumps vaccination encephalitis	Condition	SNOMED
37110324	Encephalitis caused by Venezuelan equine encephalomyelitis virus	Condition	SNOMED
4207307	Infective meningitis	Condition	SNOMED
381783	Viral infection of central nervous system	Condition	SNOMED

Table S7. Codelist of thrombocytopaenia definitions.

Concept ID	Concept name	Domain	Vocabulary
432870	Thrombocytopenic disorder	Condition	SNOMED
46272950	Thrombocytopathy, asplenia and miosis	Condition	SNOMED
44782445	Thrombocytopenia caused by alcohol	Condition	SNOMED
42536958	Pancytopenia caused by medication	Condition	SNOMED
40321716	Secondary thrombocytopenia	Condition	SNOMED
37312165	Atypical hemolytic uremic syndrome	Condition	SNOMED
37209558	Pancytopenia caused by immunosuppressant	Condition	SNOMED
37204551	Hereditary isolated aplastic anemia	Condition	SNOMED
37110394	Isolated thrombocytopenia	Condition	SNOMED
37019055	Aplastic anemia co-occurrent with human immunodeficiency virus infection	Condition	SNOMED
37018663	Thrombocytopenia co-occurrent and due to alcoholism	Condition	SNOMED
37017607	Antibody mediated acquired pure red cell aplasia caused by erythropoiesis stimulating agent	Condition	SNOMED
37016151	Aplastic anemia caused by antineoplastic agent	Condition	SNOMED
36716406	Severe fever with thrombocytopenia syndrome	Condition	SNOMED
36715586	Refractory thrombocytopenia	Condition	SNOMED
36715053	Autosomal dominant macrothrombocytopenia	Condition	SNOMED
36713112	Pancytopenia due to antineoplastic chemotherapy	Condition	SNOMED
35623407	Adult pure red cell aplasia	Condition	SNOMED
4345236	Parvoviral aplastic crisis	Condition	SNOMED
4338386	Thrombocytopenia due to non-immune destruction	Condition	SNOMED
4316372	HELLP syndrome	Condition	SNOMED
4301602	Thrombotic thrombocytopenic purpura	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4301128	Thrombocytopenia due to diminished platelet production	Condition	SNOMED
4299560	Thrombocytopenic purpura due to defective platelet production	Condition	SNOMED
4298690	Immunologic aplastic anemia	Condition	SNOMED
4292531	Thrombocytopenic purpura due to platelet consumption	Condition	SNOMED
4272928	Thrombocytopenia due to hypersplenism	Condition	SNOMED
4247776	Posttransfusion purpura	Condition	SNOMED
4239484	Acquired pancytopenia	Condition	SNOMED
4235220	Hereditary thrombocytopenic disorder	Condition	SNOMED
4234973	Chronic acquired pure red cell aplasia	Condition	SNOMED
4233407	Megakaryocytic aplasia	Condition	SNOMED
4230266	Autoimmune thrombotic thrombocytopenic purpura	Condition	SNOMED
4226905	Thrombocytopenia with AIDS (acquired immunodeficiency syndrome)	Condition	SNOMED
4225810	Aplastic anemia with AIDS (acquired immunodeficiency syndrome)	Condition	SNOMED
4219476	Thrombocytopenia due to defective platelet production	Condition	SNOMED
4218171	Uremic thrombocytopenia	Condition	SNOMED
4214947	Thrombocytopenic purpura associated with metabolic disorder	Condition	SNOMED
4211348	Aplastic anemia associated with pancreatitis	Condition	SNOMED
4204900	Acquired thrombotic thrombocytopenic purpura	Condition	SNOMED
4197574	Dilutional thrombocytopenia	Condition	SNOMED
4188208	Estren-Dameshek anemia	Condition	SNOMED
4186108	Aplastic anemia associated with metabolic alteration	Condition	SNOMED
4184758	Acquired aplastic anemia	Condition	SNOMED
4184200	Secondary aplastic anemia	Condition	SNOMED
4177177	Cellular immunologic aplastic anemia	Condition	SNOMED
4173278	Thrombocytopenia due to blood loss	Condition	SNOMED
4172008	Cyclic thrombocytopenia	Condition	SNOMED
4166754	Perinatal thrombocytopenia	Condition	SNOMED
4159966	Upshaw-Schulman syndrome	Condition	SNOMED
4159749	Idiopathic maternal thrombocytopenia	Condition	SNOMED
4159736	Radiation thrombocytopenia	Condition	SNOMED
4156233	Thrombocytopenia due to sequestration	Condition	SNOMED
4147049	Thrombocytopenia due to extracorporeal circulation	Condition	SNOMED
4146088	Aplastic anemia due to drugs	Condition	SNOMED
4145458	Thrombocytopenia due to hypothermia	Condition	SNOMED
4140545	Post infectious thrombocytopenic purpura	Condition	SNOMED
4139555	Thrombocytopenia due to massive blood transfusion	Condition	SNOMED
4137430	Idiopathic thrombocytopenic purpura	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4133984	Alloimmune thrombocytopenia	Condition	SNOMED
4133983	Secondary autoimmune thrombocytopenia	Condition	SNOMED
4133981	Benign gestational thrombocytopenia	Condition	SNOMED
4125496	Pure red cell aplasia, acquired	Condition	SNOMED
4125494	Pancytopenia with pancreatitis	Condition	SNOMED
4121264	Epstein syndrome	Condition	SNOMED
4120620	Amegakaryocytic thrombocytopenia	Condition	SNOMED
4119134	Thrombocytopenic purpura	Condition	SNOMED
4103532	Immune thrombocytopenia	Condition	SNOMED
4102469	Acute idiopathic thrombocytopenic purpura	Condition	SNOMED
4101603	Thrombocytopenia due to extracorporeal circulation of blood	Condition	SNOMED
4101583	Aplastic anemia due to infection	Condition	SNOMED
4101582	Aplastic anemia due to chronic disease	Condition	SNOMED
4100998	Aplastic anemia caused by toxic cause	Condition	SNOMED
4098148	Thrombocytopenia caused by drugs	Condition	SNOMED
4098145	Idiopathic aplastic anemia	Condition	SNOMED
4098028	Transient acquired pure red cell aplasia	Condition	SNOMED
4098027	Aplastic anemia due to radiation	Condition	SNOMED
4082738	Autoimmune pancytopenia	Condition	SNOMED
4077348	Pancytopenia-dysmelia	Condition	SNOMED
4031699	Humoral immunologic aplastic anemia	Condition	SNOMED
4028065	Primary ITP (immune thrombocytopenia)	Condition	SNOMED
4027374	Alloimmune platelet transfusion refractoriness	Condition	SNOMED
4009307	Heparin-induced thrombocytopenia with thrombosis	Condition	SNOMED
4000065	Drug-induced immune thrombocytopenia	Condition	SNOMED
436956	Evans syndrome	Condition	SNOMED
433749	Heparin-induced thrombocytopenia	Condition	SNOMED
432881	Pancytopenia	Condition	SNOMED
318397	Chronic idiopathic thrombocytopenic purpura	Condition	SNOMED
140681	Constitutional aplastic anemia	Condition	SNOMED
137829	Aplastic anemia	Condition	SNOMED
138723	Acquired red cell aplasia	Condition	SNOMED
432870	Thrombocytopenic disorder	Condition	SNOMED
432881	Pancytopenia	Condition	SNOMED
433749	Heparin-induced thrombocytopenia	Condition	SNOMED
436956	Evans syndrome	Condition	SNOMED
440372	Acquired thrombocytopenia	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
441264	Primary thrombocytopenia	Condition	SNOMED
4000065	Drug-induced immune thrombocytopenia	Condition	SNOMED
4009307	Heparin-induced thrombocytopenia with thrombosis	Condition	SNOMED
4027374	Alloimmune platelet transfusion refractoriness	Condition	SNOMED
4028065	Primary ITP (immune thrombocytopenia)	Condition	SNOMED
4031699	Humoral immunologic aplastic anemia	Condition	SNOMED
4077348	Pancytopenia-dysmelia	Condition	SNOMED
4082738	Autoimmune pancytopenia	Condition	SNOMED
4098027	Aplastic anemia due to radiation	Condition	SNOMED
4098028	Transient acquired pure red cell aplasia	Condition	SNOMED
4098145	Idiopathic aplastic anemia	Condition	SNOMED
4098148	Thrombocytopenia caused by drugs	Condition	SNOMED
4100998	Aplastic anemia caused by toxic cause	Condition	SNOMED
4101582	Aplastic anemia due to chronic disease	Condition	SNOMED
4101583	Aplastic anemia due to infection	Condition	SNOMED
4101603	Thrombocytopenia due to extracorporeal circulation of blood	Condition	SNOMED
4102469	Acute idiopathic thrombocytopenic purpura	Condition	SNOMED
4103532	Immune thrombocytopenia	Condition	SNOMED
4119134	Thrombocytopenic purpura	Condition	SNOMED
4120620	Amegakaryocytic thrombocytopenia	Condition	SNOMED
4121264	Epstein syndrome	Condition	SNOMED
4125494	Pancytopenia with pancreatitis	Condition	SNOMED
4125496	Pure red cell aplasia, acquired	Condition	SNOMED
4133981	Benign gestational thrombocytopenia	Condition	SNOMED
4133983	Secondary autoimmune thrombocytopenia	Condition	SNOMED
4133984	Alloimmune thrombocytopenia	Condition	SNOMED
4137430	Idiopathic thrombocytopenic purpura	Condition	SNOMED
4139555	Thrombocytopenia due to massive blood transfusion	Condition	SNOMED
4140545	Post infectious thrombocytopenic purpura	Condition	SNOMED
4145458	Thrombocytopenia due to hypothermia	Condition	SNOMED
4146088	Aplastic anemia due to drugs	Condition	SNOMED
4147049	Thrombocytopenia due to extracorporeal circulation	Condition	SNOMED
4156233	Thrombocytopenia due to sequestration	Condition	SNOMED
4159736	Radiation thrombocytopenia	Condition	SNOMED
4159749	Idiopathic maternal thrombocytopenia	Condition	SNOMED
4159966	Upshaw-Schulman syndrome	Condition	SNOMED
4166754	Perinatal thrombocytopenia	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
4172008	Cyclic thrombocytopenia	Condition	SNOMED
4173278	Thrombocytopenia due to blood loss	Condition	SNOMED
4177177	Cellular immunologic aplastic anemia	Condition	SNOMED
4184200	Secondary aplastic anemia	Condition	SNOMED
4184758	Acquired aplastic anemia	Condition	SNOMED
4186108	Aplastic anemia associated with metabolic alteration	Condition	SNOMED
4188208	Estren-Dameshek anemia	Condition	SNOMED
4197574	Dilutional thrombocytopenia	Condition	SNOMED
4204900	Acquired thrombotic thrombocytopenic purpura	Condition	SNOMED
4211348	Aplastic anemia associated with pancreatitis	Condition	SNOMED
4214947	Thrombocytopenic purpura associated with metabolic disorder	Condition	SNOMED
4218171	Uremic thrombocytopenia	Condition	SNOMED
4219476	Thrombocytopenia due to defective platelet production	Condition	SNOMED
4225810	Aplastic anemia with AIDS (acquired immunodeficiency syndrome)	Condition	SNOMED
4226905	Thrombocytopenia with AIDS (acquired immunodeficiency syndrome)	Condition	SNOMED
4230266	Autoimmune thrombotic thrombocytopenic purpura	Condition	SNOMED
4233407	Megakaryocytic aplasia	Condition	SNOMED
4234973	Chronic acquired pure red cell aplasia	Condition	SNOMED
4235220	Hereditary thrombocytopenic disorder	Condition	SNOMED
4239484	Acquired pancytopenia	Condition	SNOMED
4247776	Posttransfusion purpura	Condition	SNOMED
4272928	Thrombocytopenia due to hypersplenism	Condition	SNOMED
4292531	Thrombocytopenic purpura due to platelet consumption	Condition	SNOMED
4298690	Immunologic aplastic anemia	Condition	SNOMED
4299560	Thrombocytopenic purpura due to defective platelet production	Condition	SNOMED
4301128	Thrombocytopenia due to diminished platelet production	Condition	SNOMED
4301602	Thrombotic thrombocytopenic purpura	Condition	SNOMED
4316372	HELLP syndrome	Condition	SNOMED
4338386	Thrombocytopenia due to non-immune destruction	Condition	SNOMED
4345236	Parvoviral aplastic crisis	Condition	SNOMED
35623407	Adult pure red cell aplasia	Condition	SNOMED
36713112	Pancytopenia due to antineoplastic chemotherapy	Condition	SNOMED
36715053	Autosomal dominant macrothrombocytopenia	Condition	SNOMED
36715586	Refractory thrombocytopenia	Condition	SNOMED
36716406	Severe fever with thrombocytopenia syndrome	Condition	SNOMED
37016151	Aplastic anemia caused by antineoplastic agent	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
37017607	Antibody mediated acquired pure red cell aplasia caused by erythropoiesis stimulating agent	Condition	SNOMED
37018663	Thrombocytopenia co-occurrent and due to alcoholism	Condition	SNOMED
37019055	Aplastic anemia co-occurrent with human immunodeficiency virus infection	Condition	SNOMED
37110394	Isolated thrombocytopenia	Condition	SNOMED
37204551	Hereditary isolated aplastic anemia	Condition	SNOMED
37209558	Pancytopenia caused by immunosuppressant	Condition	SNOMED
37312165	Atypical hemolytic uremic syndrome	Condition	SNOMED
40321716	Secondary thrombocytopenia	Condition	SNOMED
42536958	Pancytopenia caused by medication	Condition	SNOMED
44782445	Thrombocytopenia caused by alcohol	Condition	SNOMED
46272950	Thrombocytopathy, asplenia and miosis	Condition	SNOMED
440982	Wiskott-Aldrich syndrome	Condition	SNOMED
435076	Transient neonatal thrombocytopenia	Condition	SNOMED
4345345	Neonatal alloimmune thrombocytopenia	Condition	SNOMED
4311682	Radial aplasia-thrombocytopenia syndrome	Condition	SNOMED
37393863	Platelet count	Measurement	SNOMED
401296001	Percentage reticulated platelet count	Measurement	SNOMED
4267147	Platelet count	Measurement	SNOMED
4260199	Serial platelet counts	Measurement	SNOMED
4038124	Platelet count, Rees-Ecker method	Measurement	SNOMED
4016231	Platelet count, blood, manual	Measurement	SNOMED
4015170	Platelet count, blood, automated	Measurement	SNOMED
3039827	Platelets [#/volume] in Body fluid by Automated count	Measurement	SNOMED
3031586	Platelets [#/volume] in Blood by Estimate	Measurement	SNOMED
3024386	Platelet mean volume [Entitic volume] in Blood by Rees-Ecker	Measurement	SNOMED
3007461	Platelets [#/volume] in Blood	Measurement	SNOMED
3007461	Platelets [#/volume] in Blood	Measurement	SNOMED
3024386	Platelet [Entitic mean volume] in Blood by Rees-Ecker	Measurement	SNOMED
3031586	Platelets [#/volume] in Blood by Estimate	Measurement	SNOMED
3039827	Platelets [#/volume] in Body fluid by Automated count	Measurement	SNOMED
4015170	Platelet count, blood, automated	Measurement	SNOMED
4016231	Platelet count, blood, manual	Measurement	SNOMED
4038124	Platelet count, Rees-Ecker method	Measurement	SNOMED
4260199	Serial platelet counts	Measurement	SNOMED
4267147	Platelet count	Measurement	SNOMED
4270019	Percentage reticulated platelet count	Measurement	SNOMED

Concept ID	Concept name	Domain	Vocabulary
3024929	Platelets [#/volume] in Blood by Automated count	Measurement	SNOMED

ANNEX V. Glossary

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

Aggregated Data

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

Benefit-Risk Assessment

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

Common Data Model (CDM)

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

Complex Studies (C3)

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

Coordination Centre (CC)

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

Data Access

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

Data Quality Framework

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

Data Source

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

DARWIN EU®

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

EMA (European Medicines Agency)

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

Evidence Generation

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

Federated Network

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

GDPR (General Data Protection Regulation)

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

Health Technology Assessment (HTA)

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

Metadata

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

Off-the-Shelf Studies (OTS)

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

OHDSI (Observational Health Data Sciences and Informatics)

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

Patient-Level Data

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

OMOP (Observational Medical Outcomes Partnership)

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

Real-World Data (RWD)

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

Real-World Evidence (RWE)

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

Regulatory Decision-Making

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

Routine Repeated Studies (RR)

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

Study Protocol

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

Very Complex Studies (C4)

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