Study Protocol

P4-C2-018

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

10/03/2026

Version 3.0

Authors: Ivan Lam, Daniel Prieto-Alhambra

Confidential

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
Protocol version	V3.0
Date	10/03/2026
EUPAS number	EUPAS1000000969
Active Substance	N/A
Medicinal Product	N/A
Research question and objectives	The aim of this study is to estimate the background incidence rates of selected vaccine adverse events of special interest in Europe. 1. To estimate population level incidence rates of selected adverse events of special interest (AESIs) in the general population during 2015 and until the latest data availability, stratified by calendar period based on COVID-19 pre-, during- and post- pandemic periods, calendar quarter, sex, age groups, sex and age groups, and data source. 2. To estimate age and sex standardised incidence rates (to the European population) of the selected AESIs in the general population during 2015 and until the latest data availability. 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, Daniel Prieto-Alhambra

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

LIST OF ABBREVIATIONS

Acronyms/terms	Description
AESI	Adverse events of special interest
ANVISA	Brazilian Health Regulatory Agency
BGR	Background incidence rates
CC	Coordination Centre
CDM	Common Data Model
CPRD	Clinical Practice Research Datalink
COVID-19	Coronavirus disease-2019
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
EHR	Electronic Health Record
EMA	The European Medical Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ESP	European Standard Population
FinOMOP-THL	Finnish Care Register for Health Care
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
IPCI	Integrated Primary Care Information
MHRA	Medicines and Healthcare products Regulatory Agency
NLHR	Norwegian Linked Health Registry data
OMOP	Observational Medical Outcomes Partnership
PHE	Public Health Emergencies
RT-PCR	Reverse transcription polymerase chain reaction
RWD	Real-World Data
RWE	Real-World Evidence
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SIDIAP	The Information System for Research in Primary Care
SNOMED	Systematized Nomenclature of Medicine
SPEAC	Safety Platform for Emergency vACcines
WG	Working Group
WHO	World Health Organisation

1. TITLE

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

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigators	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)	Gustav Klingstedt Tiina Wahlfors Toni Lehtonen	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 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 on Primary Care (SIDIAP)	Elena Roel Herranz Anna Palomar Cros Augustina Giuliodori Picco Laura Granés	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

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 vaccine safety concerns is crucial for maintaining public confidence. Information on background incidence rates of adverse events of special interest (AESIs) can support this response when used for rapid assessment via observed-to-expected analyses, an essential first step to inform further signal management. While evidence from real-world data (RWD) studies has been published during and after the pandemic, background rates vary across countries, 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 International Coalition of Medicines Regulatory Authorities (ICMRA) Working Group on Real-World Evidence for Public Health Emergencies was established to provide a collaborative forum for international regulatory agencies to proactively enhance the efficiency and coordination of critical responses to emerging public health emergencies (PHEs) through joint evidence generation. This study aims to support a proof-of-concept within the Working Group, with the goal to generate evidence on background rates of eight vaccine AESIs selected based on capacity and interest at national level. Parallel studies will be conducted by the following jurisdictions in collaboration with local research teams: Brazilian Health Regulatory Agency (ANVISA; Brazil), Health Canada, Health Sciences Authority (HSA; Singapore), and the Medicines and Healthcare products Regulatory Agency (MHRA; UK). A secondary goal of the group will be to critically review existing phenotypes to improve (as 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, building on the previous DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe, will also support vaccine safety monitoring endeavours at European level as part of the Vaccine Monitoring Platform.

Objectives

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

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

- Descriptive disease epidemiology study to estimate crude population level incidence rates, age, and sex standardised incidence rates (to the European population) of selected adverse events of special interest (AESIs) in the general population during 2015 and until the latest data availability, stratified by calendar period based on pre- (2015-2019), during (2020-2022) and post-COVID-19 pandemic periods (2023 to latest available data), sex, age groups, and data source (Objectives 1 and 2)

Population

The study population will include all individuals observed in one of the participating data sources during the study period. We will require individuals to have at least 365 days of data availability before entering the cohort. The index date of cohort entry will be 1st January 2015 or the date that individuals fulfil the data availability. The individuals have to have no prior record of that specific outcome within a 90-day 'clean window' period.

Variables

Exposure:

Not applicable

Outcome:

The outcomes of this study (AESIs) are: Guillain-Barre Syndrome (GBS), myocarditis (+/- pericarditis), encephalitis, stroke (ischaemic stroke and haemorrhagic stroke), thrombocytopenia, and immune thrombocytopenia.

Relevant covariates:

The relevant covariates include 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) and calendar year based on pre- (2015-2019), during (2020-2022) and post-COVID-19 pandemic periods (2023 to latest available data)

Data sources

1. Denmark: Danish Data Health Registries (DK-DHR)
2. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
3. Germany: InGef Research Database (InGef RDB)
4. The Netherlands: Integrated Primary Care Information (IPCI)
5. Norway: Norwegian Linked Health Registry data (NLHR)
6. Spain: The Information System for Research on Primary Care (SIDIAP)
7. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
8. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Four data sources including DK-DHR, FinOMOP-THL, NLHR and HI-SPEED consist of nationwide registry data. Two data sources (IPCI and CPRD GOLD) consist of primary care data.

Statistical Analysis

The phenotypes for the study outcomes have been defined in the previous study following the Standard Operating Procedure under a dynamic workflow between the study team and the EMA.

Incidence rates per 100,000 person-years and corresponding 95% confidence intervals will be reported using the exact Poisson model. Analysis will be stratified by calendar month based on pre-, during and post-COVID-19 pandemic periods, year, age group, and sex within each database. Incidence rates will not be estimated if there are less than 5 events in a given stratum. We will also standardise these rates to the European population based on the European Standard Population 2013 for Objective 2. Meta-analyses of

incidence rates of AESIs will be conducted (with 95% prediction intervals) to pool estimates from data sources grouped according to healthcare setting.

4. AMENDMENTS AND UPDATES

None

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	January 2026
Creation of Analytical code	February 2026
Execution of Analytical Code on the data	March 2026
Draft Study Report	April 2026
Final Study Report	May 2026

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

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 incidence 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] The heterogeneity can come from different clinical coding systems, health care delivery systems, clinical practice, or reflect the true differences between 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 the demographic and clinical characteristics of patients could provide useful information for evaluating potential 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, building on the former COVID-19 Real-World Evidence (RWE) and Observational Studies Working Group,[9] thus addressing the need for collaboration beyond COVID-19 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 aims to support a proof-of-concept with the ultimate goal of generating evidence on background rates of several selected vaccine adverse events of special interest (AESIs) within the WG. Parallel studies will be conducted in the following jurisdictions in collaboration with local research teams: Brazilian Health Regulatory Agency (ANVISA; Brazil), The European Medical Agency (EMA), Health Canada, Health Sciences Authority (HSA; Singapore), and Medicines and Healthcare products Regulatory Agency (MHRA; UK). A secondary goal of the WG will be to critically review existing phenotypes to improve (as 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 will support vaccine safety monitoring endeavours as part of the Vaccine Monitoring platform.[10]

This study, building on previous the DARWIN EU® study (EUPAS1000000254) proposed by the EMA, focused on estimating the background incidence rates of a shortlist of AESIs selected by the WG at European level as part of the Vaccines Monitoring Platform. These AESIs represent current safety concerns at regional level and are considered a priori 'feasible' by the WG. This study is now being repeated to generate evidence on background incidence rates of AESIs for vaccines with greater granularity of the data, in order to best support future observed-to-expected analyses for specific vaccine products.

7. RESEARCH QUESTION AND OBJECTIVES

Research questions

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

Research objectives

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

The specific objectives of this study are:

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

2. To estimate age and sex standardised incidence rates (to the European population) of the selected AESIs in the general population during 2015 and until the latest data availability, stratified by calendar period based on COVID-19 pre-, during- and post-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 will be conducted using routinely collected health data from eight data sources from eight countries across Europe including seven EU Member States. The study will comprise of:

- Population-level descriptive epidemiology analyses will be conducted to address objectives 1 and 2, assessing population level, sex, and age standardised incidence rates of selected AESIs.

8.2. Follow-up

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

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

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

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

Individuals will be followed until the earliest date of the study events of interest, death, exceeding specified age range (in age-group specific analysis), end of observation period in the database, or end of data availability of data source. **(Table 1)**

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

Table 1. Operational Definition of Time 0 (index date) and other primary time anchors.

Study population name(s)	Time Anchor Description (e.g. time 0)	Number of entries	Type of entry	Washout window	Care Setting [1]	Incident with respect to…
General population	a) study start date (1st January 2015), b) date at which they have sufficient prior data availability (365 days), c) date on which they fulfil the 'clean window' criteria of a specific event.	multi	General population	n/a	IP, OP	n/a
Individuals with observed AESIs	Date of diagnosis	multi	Incident	90 days	IP, OP	No record of the same condition in the wash out period.

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable, AESI = Adverse events of special interest.

8.3. Study population with inclusion and exclusion criteria

General population:

The study population will include all individuals observed in one of the participating data sources during the study period. We will require individuals to have at least 365 days of data availability before entering the cohort. The index date of cohort entry will be 1st January 2015 or the date that individual fulfil the data availability and outcome 'clean window' (details **Section 8.6.2**) requirement.

8.4. Study setting and data sources

This study will be conducted using routinely collected data from eight primary, secondary care, and hospital data sources in the DARWIN EU® network of data partners from eight European countries. Four data sources, including DK-DHR, FinOMOP-THL, NLHR and HI-SPEED, consist of nationwide registry data. Two data sources, including IPCI and CPRD GOLD, consist of primary care only data. All data were *a priori* mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). **(Table 2)**

Table 2. Data sources.

Country	Name of Data source[2]	Health Care setting[1]	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to
Denmark	Danish Data Health Registries (DK-DHR)	Community pharmacies, secondary care – specialists, hospital inpatient care	Registries	5.9M	1st January 2015 to 04th November 2025	All objectives
Finland	Finnish Care Register for Health Care (FinOMOP-THL)	Primary care, secondary care – specialists (ambulatory or hospital OP care), hospital IP care	HER, Registries	5.7M	1st January 2015 to 11th November 2025	All objectives
Germany	Institute for Applied Health Research Berlin GmbH (InGef)	Primary care – GPs, community pharmacists, primary care specialists, secondary care – specialists, hospital inpatient care	Claims	7.6M	1st January 2016 to 05th November 2025	All objectives
The Netherlands	Integrated Primary Care Information (IPCI)	Primary care	EHR	1.33M	1st January 2015 to 15th October 2025	All objectives
Norway	Norwegian Linked Health Registry (NLHR)	Primary care, secondary care – specialists (ambulatory or hospital OP	Registries	5.55M	1st January 2015 to 21st August 2025	All objectives

		care), hospital IP care				
Spain	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP)	Primary care database + linkage to hospital data	EHR	5.9M	1st January 2015 to 05th November 2025	All objectives
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 2015 to 31st August 2025	All objectives
The United Kingdom	Clinical Practice Research Datalink (CPRD) GOLD	Primary care	EHR	2.6M	1st January 2015 to 20th May 2025	All objectives

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable, EHR = Electronic Health Record
[2] DK-DHR = Danish Data Health Registries, FinOMOP-THL = Finnish Care Register for Health Care, InGef = Institute for Applied Health Research Berlin GmbH, IPCI = Integrated Primary Care Information, NLHR = Norwegian Linked Health Registry, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària, HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD = Clinical Practice Research Datalink GOLD

[3] Primary care data from HI-SPEED only covers two regions in Sweden (Västra Götaland region and Stockholm region), amounting to about 40% of the total population.

Data sources selection

These data sources fulfil 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 is from 01 January 2015 to the most recent data available for each contributing data source. For stratification purposes, the period will be split into: 2015-2019, 2020-2022 and 2023 to latest data availability, taking into account the impact of the COVID-19 pandemic.

8.6. Variables

8.6.1. Outcome

AESIs:

The AESIs of interest are listed below. They were selected by the ICMRA study team, based on national capacity and interest, while aligning across jurisdictions. The phenotypes have been defined in the previous study following the Standard Operating Procedure under a dynamic workflow between the study team and the EMA. In principle, broad and narrow phenotype definitions will be 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 will be defined as having diagnosis records for both conditions on the same index date, whilst incidence of myocarditis without pericarditis will be defined as the absence of diagnosis record of pericarditis on the same index day as the diagnosis of myocarditis.

For each study outcome, a clean window will be applied to define incident outcomes. We will apply an event specific clean window as shown in **Table 3** for acute and recurrent outcomes.

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

Table 3. Summary of event specific clean windows for outcomes

	Outcome	Event specific clean window**	Type
Immune-mediated diseases	Guillain Barré syndrome	90	Acute
Blood disorders	Thrombocytopenia	90	Acute
	Immune Thrombocytopenia	90	Acute
Cardiovascular system	Myocarditis (+/-Pericarditis)	90	Acute
	Stroke	90	Acute
	Ischaemic stroke	90	Acute
	Haemorrhagic stroke	90	Acute
Nervous system	Encephalitis	90	Acute

**Event specific clean window: For each study outcome, a specific clean window will be applied to define incident outcomes. This definition is subject to changes during the phenotype stage.

The preliminary concept sets used for the identification of outcomes are described in **Annex IV**.

Sensitivity analyses using a narrow (more specific) case definition for certain outcomes and extended the clean windows of 365 days described in **Table 5** will be conducted.

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

Demographics:

Sex: Female, male.

Age

Age groups:

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

Given that some study events may be very rare and have counts below 5 under the five and 10-years age groups, a more broader age group will be used as well: 0 – 18, 19 – 64, 65+.

8.7. Study size

The population base for this study will be approximately 73.2M, estimated based on the "total number of patients" from all the included data sources as reported in the DARWIN EU® Portal. This does not reflect the actual number of individuals in each year during the study period, as the numbers of individuals in each data source is dynamic.

8.8. Analysis

8.8.1. Federated network analyses

All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing individuals' data.

Before sharing the study package, test runs of the analytics will be performed on a subset of the data sources, and quality control checks will be performed. After all the tests are passed (see **Annex III.**), the final package will be released in a version-controlled study repository for execution against all the participating data sources.

8.8.2. Data privacy protection

The data partners will locally execute the analytics against the OMOP CDM in R Studio and review and approve the default aggregated results. They will then be made available to the Principal Investigators and study team in secure online repository of DTZ (Data Transfer Zone). All results will be locked and timestamped for reproducibility and transparency. The study results of all data sources will be checked, after which they are made available to the team, and the Study Dissemination Phase can start. All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing patient-level data. Cell counts <5 will be suppressed when reporting results to comply with the data source's privacy protection regulations.

8.8.3. Statistical model specification and assumptions of the analytical approach considered

Objectives 1 and 2

Incidence rates per 100,000 person-years will be calculated as the number of incident cases divided by the total person-time at risk. The corresponding 95% confidence intervals will be reported using exact Poisson model. A pre-defined clean window of 90 days will be applied to each AESI, an individual can re-enter the cohort after the clean window of the previous event been fulfilled. **(Table 4)**

The incidence rates will be calculated using the "*IncidencePrevalence*" R package, developed by DARWIN EU®.[11]

Stratification/subgroup

- By database
- By calendar quarter
- By calendar year based on pre- (2015-2019), during (2020-2022) and post-COVID-19 pandemic periods (2023 to latest available data)
- By age group stated in **8.6 Variables**. A wider age will be used if most of the subgroups have event count less than 5.
- By sex
- By age group - sex

Incidence rates will not be estimated if there are less than 5 events in a given stratum.

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

Table 4. 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 8.6.2 Outcomes**[1]
Analytic software:	R
Model: ***(provide details or code)***	We will estimate incidence rates and 95% confidence interval. Overall, age group, and sex specific rates will be reported when event number is larger than 5 within the strata.
Confounding adjustment method	***Name method and provide relevant details, e.g. bivariate, multivariable, propensity score matching (specify matching algorithm ratio and calliper), propensity score weighting (specify weight formula, trimming, truncation), propensity score stratification (specify strata definition), other.***
	For incidence rates, we will estimate the rates with stratification and standardisation
Subgroup Analyses	***List all subgroups***
	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-analysis of results based on a random effect model will be conducted to synthesise pooled findings across data sources stratified by the following healthcare settings and data types: 1) nationwide registry data: DK-DHR, FinOMOP-THL, NLHR and HI-SPEED; 2) primary care only data sources: IPCI and CPRD GOLD. Meta-analytic estimates of incidence of AESI with 95% prediction intervals will be computed.

Sensitivity analyses

Description of sensitivity analyses are presented by means of **Table 5**.

Table 5. 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 narrower case definition, tailored to each event and clinically relevant will be defined during the phenotyping process, then applied in the sensitivity analysis. The outcomes for which a narrower case definition will be applied include: Encephalitis, Guillain Barré syndrome, Stroke, Ischaemic stroke, Haemorrhagic stroke, and Immune thrombocytopenia	The narrower definition of the outcome may impact on the included study population and outcome cases. If the case definition is narrower, we might see lower incidence rates for the specific event.	A narrower case definition may increase the specificity of incident event	The narrower case definition for outcome may result in under capturing of incident cases of specific outcome
Extended clean window	A longer clean window of 365 days for all outcomes will be implemented to define incident cases of outcomes	A longer clean window may impact 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 will increase the specificity of the inclusion of incident events	A longer clean window may lead to under capturing incident cases of a specific outcome

8.8.4. Output

Output will include the following:

A PDF report including an executive summary, and the following table(s) and figure(s).

- Table 7. *Cohort counts for Individuals with AESIs*
- Figure 1. *Crude incidence rates per outcome stratified by database over full study period.*
- Figure 2. *Standardised incidence rates per outcome stratified by database over full study period.*
- Figure 3. *Incidence rates per outcome stratified by age groups.*
- Figure 4. *Incidence rates per outcome stratified by sex.*
- Figure 5. Incidence rates *per outcome* stratified by calendar year and per pre/during/post pandemic
- Figure 6. Pooled incidence rates of AESIs by healthcare settings.

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

Table 6. Cohort counts for Individuals with AESIs

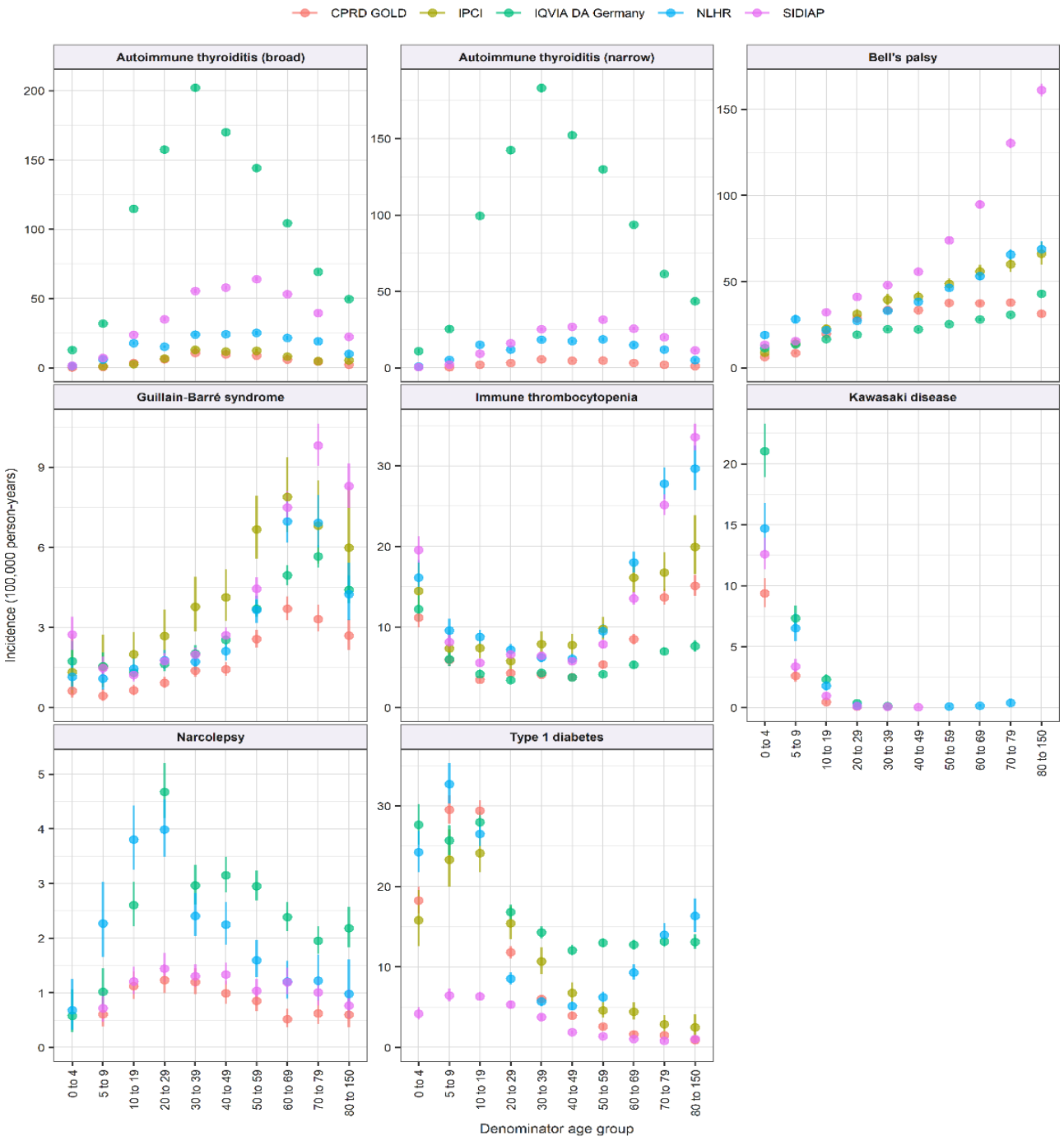
	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5
(AESIs)					
Number subjects					
Age (Median [IQR])					
Sex, male (n (%))					

Outcomes classified as very rare if the maximum incidence rate was below 10 per 100,000 person-years, rare if between 10 and 100 per 100,000 person-years, or uncommon to common otherwise. The data presented are included for illustration purpose only.

Figure 1. Example of crude incidence rates per outcome stratified by database over full study period.

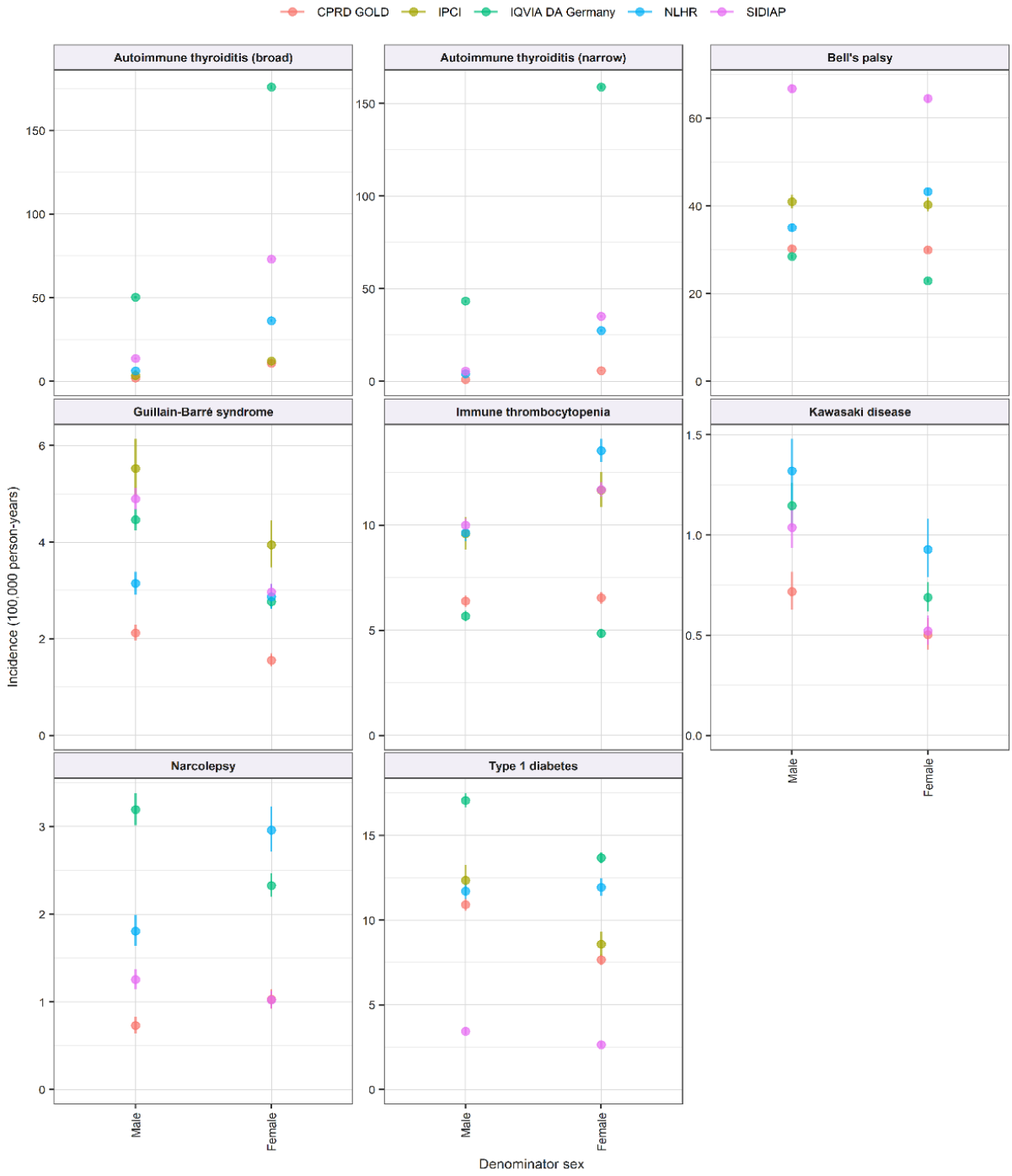
Outcomes classified as very rare if the maximum incidence rate was below 10 per 100,000 person-years, rare if between 10 and 100 per 100,000 person-years, or uncommon to common otherwise. The data presented are included for illustration purpose only.

Figure 2. Example of standardised incidence rates per outcome stratified by database over full study period.



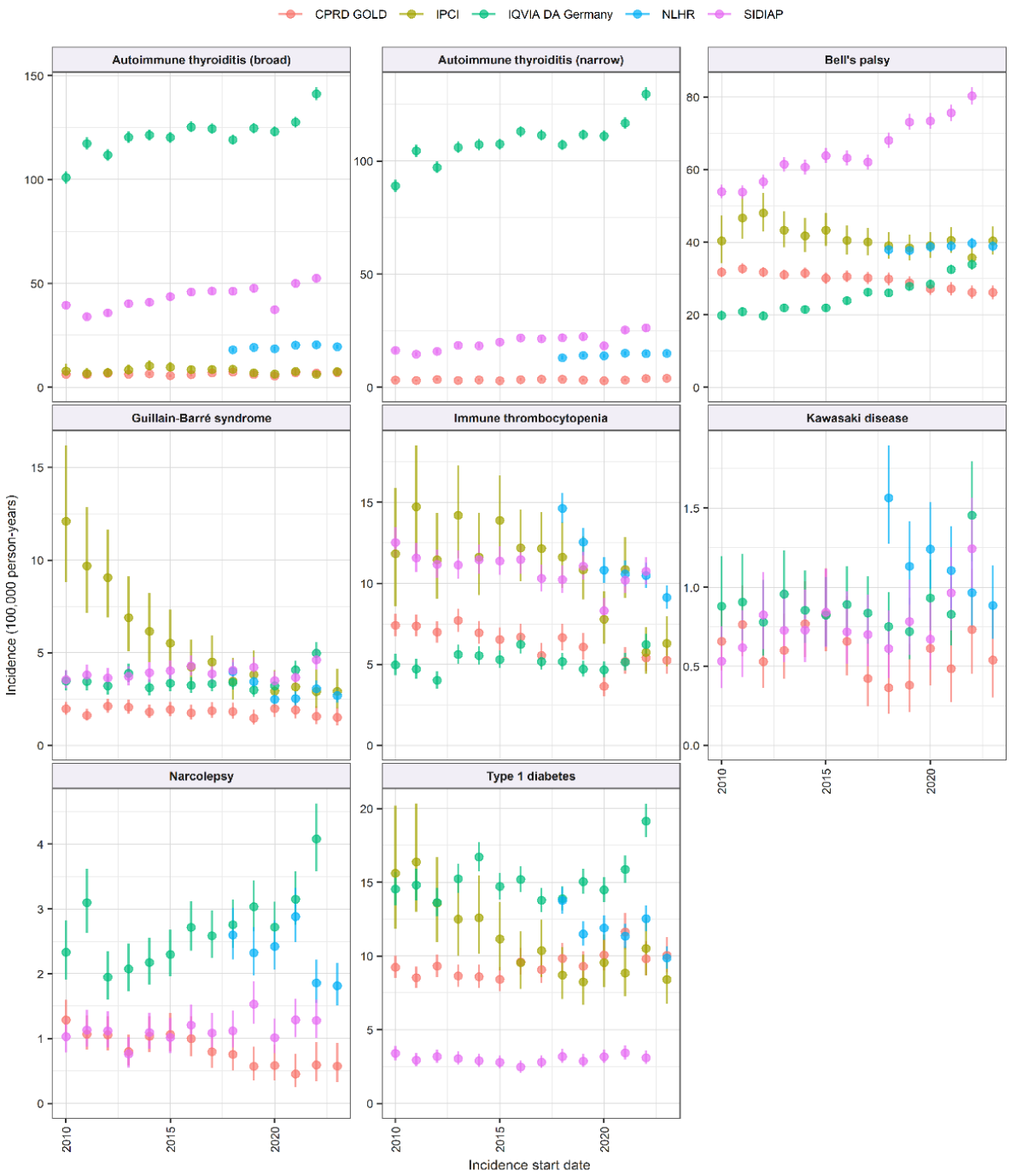
The data presented are included for illustration purpose only.

Figure 3. Example of incidence rates *per outcome* stratified by age groups.



The data presented are included for illustration purpose only.

Figure 4. Example of incidence rates of *per outcome* stratified by sex.



The data presented are included for illustration purpose only.

Figure 5. Example of incidence rates *per outcome* stratified by calendar quarter and year.

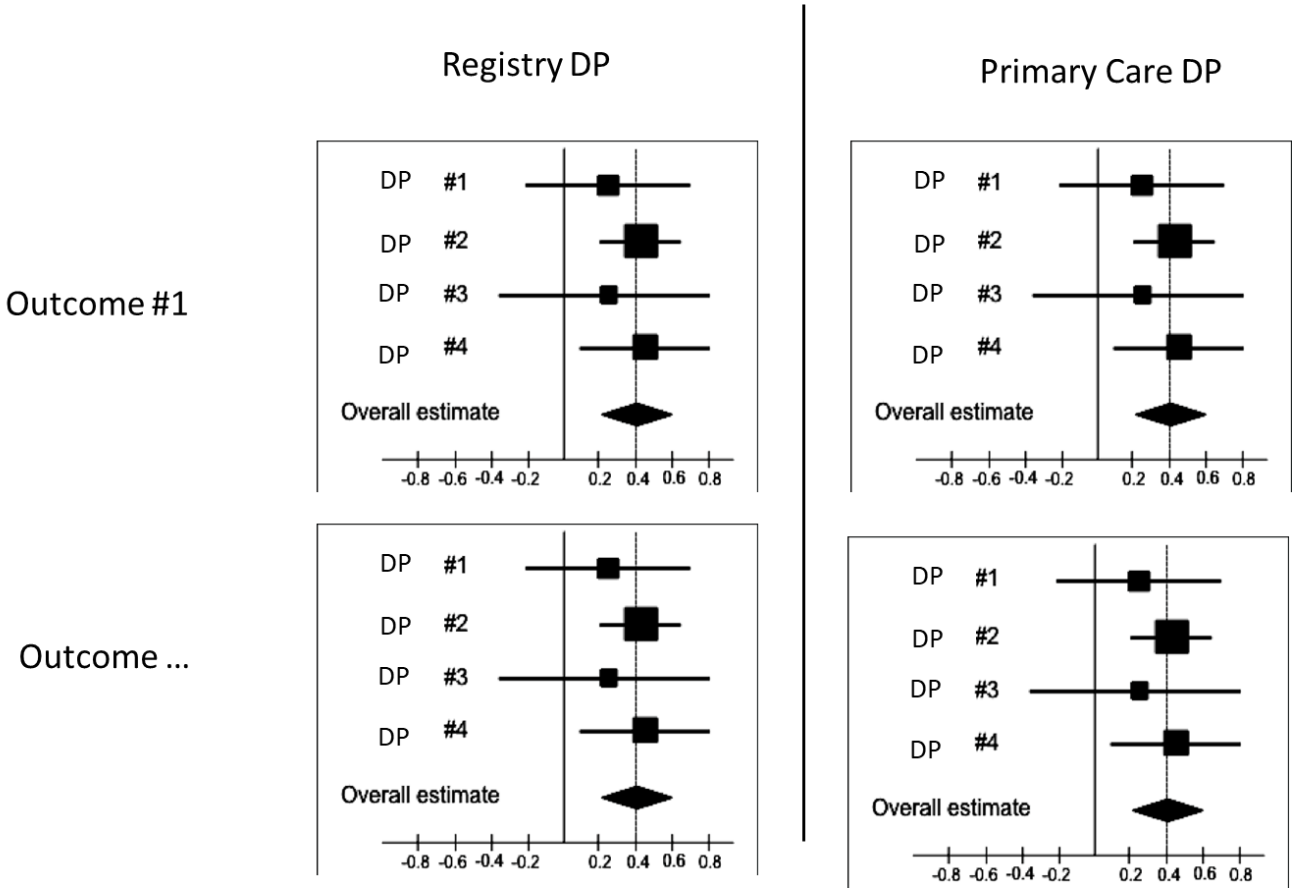


Figure 6. Example of pooled incidence rates of AESI in registry and primary care only databases.

8.9. Evidence synthesis

Results from analyses described in **Section 8.8** will be presented separately for each data source. Meta-analysis of results based on a random effect model will be conducted to synthesise findings across 1) nationwide registry data; 2) primary care only data sources.

9. STRENGTHS AND LIMITATIONS

General limitations:

The results estimated from this study will only reflect the populations from the included data sources. Electronic health records have certain inherent limitations because they were collected for clinical purpose rather than primarily for research use.

We assume that if there were no related clinical codes of a condition presented for an individual in the data, the condition does not exist for this individual.

Misclassification of outcomes could happen if individuals received health care service outside of the data capture system. For example, in the UK primary care data, we would not be able to capture events that had been recorded in private care sectors. However, all the selected data sources are representative of the general source population, and the potential impact of misclassification is expected to be similar within the data source across the study period. The data source setting will also impact capture of diagnosis, therefore data sources that record only primary diagnoses such as IPCI and CPRD GOLD will have a lower number of cases if the diagnosis is mostly made in hospital, which is the case for most of the studied outcomes.

Study-specific limitations:

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

Since published literature is not available for all AESIs to determine the appropriate length of the clean period, it is possible that some periods were set incorrectly and may lead some rates to reflect a combination of prevalent and incident cases, especially for chronic conditions. The sensitivity analyses are designed to explore this variation.

Additionally, changes in clinical guidelines or practice of recording of the electronic health records could affect the estimation of incidence rates over time. Primary care data from HI-SPEED only covers two regions in Sweden (Västra Götaland region and Stockholm region), amounting to about 40% of the total population. Abnormal platelet measurement records used to identify incidence of thrombocytopenia may be unavailable in certain databases including FinOMOP-THL, HI-SPEED and InGef.

10. 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/.

11. ANNEXES

ANNEX I. Description of data sources

DATA SOURCES DESCRIPTION

#	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

#	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

#	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/

#	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

#	Section	Description
1	Data source identification and country	NLHR (Norwegian Linked Health Registry data) Norway

#	Section	Description
2	Data partner information section	University of Oslo Faculty of Mathematics and Natural Science – Department of Pharmacy
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/

#	Section	Description
1	Database Identification and country	HI-SPEED (Health Impact - Swedish Population Evidence Enabling Data-linkage) Sweden
2	Data partner information section	SMPA-GU, Läkemedelsverket, Box 26 Pharmacoepidemiology and Analysis Department (FeA)
3	Coverage and timespan	Data collection since: 2020 Extent: Nation-wide. The catchment area includes the whole of Sweden, covering the full population of approximately 10 million.
4	Healthcare setting / type of data	Primary care – gps, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: Socio-demographics, drug use and prescriptions, diagnoses, cause of death, primary care procedures and visits, as well as secondary care and inpatient visits or clinical events.
5	Data collection process	Registries. The data is acquired from the Swedish nation 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
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, ICD10 is used for diagnoses, and the Swedish procedure coding system (KVA) is used for clincal 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. No.
9	Quality control (database specific)	The source data is obtained from the Swedish National and Regional Registers. The registers perform some regular quality controls on their data. After receiving the data, we 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.
11	Vital status	Data on death and cause-of-death are extracted from the Cause-of-Death registry.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: Website:

#	Section	Description
1	Data source identification and country	SIDIAP (The Information System for the Development of Research in Primary Care) Catalunya, Spain
2	Data partner information section	IDIAPJGol
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. SIDIAP is a database of primary care electronic health records of the population of Catalonia,

#	Section	Description
		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.
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.

#	Section	Description
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

#	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 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.

#	Section	Description
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 will be included in this study because it is a secondary care and hospital data source that provides 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 will be 1.41M.

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

There are no study specific limitations present in DK-DHR.
Lastly, DK-DHR has blanket approval, which makes the execution of this study feasible within the current study timelines.

Finnish Care Register for Health Care (FinOMOP-THL)

FinOMOP-THL will be included in this study because it is a primary care and hospital data source that provides 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 will be 968.3k.

Moreover, data availability and follow-up in FinOMOP-THL is sufficient, as data availability starts in 2011, and the date of most recent data extraction is 01 October 2024, which aligns with the study period. The median follow-up of the first observation period in FinOMOP-THL is 5,020 days.

. There are 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 is estimated to take a week from submission, which makes the execution of this study feasible within the current study timelines.

InGef Research Database (InGef RDB)

InGef RDB will be included in this study because it is a claims data source that provides 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 will be 1.22M.

Moreover, data availability and follow-up in InGef RDB is sufficient, as data availability starts in 2016, and the date of most recent data extraction is 02 Jan 2026, which aligns with the study period. The median follow-up of the first observation period in InGef RDB is 3,560 days.

There are 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 will be 01 January 2016 to study period end date.

Lastly, IRB approval for InGef RDB is estimated to take 2-4 weeks, which makes the execution of this study feasible within the current study timelines.

Integrated Primary Care Information (IPCI)

IPCI will be included in this study because it is a primary care data source that provides 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 will be 236.8k.

Moreover, data availability and follow-up in IPCI is sufficient, as data availability starts in 2006, and the date of most recent data extraction is 01 December 2024, which aligns with the study period. The median follow-up of the first observation period in IPCI is 1,730 days.

There are 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 has blanket approval, which makes the execution of this study feasible within the current study timelines.

Norwegian Linked Health Registry (NLHR)

NLHR will be included in this study because it is a registry data source that provides 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 will be 720.5k.

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

There are no study specific limitations present in NLHR.
Lastly, NLHR has blanket approval, which makes the execution of this study feasible within the current study timelines.

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

HI-SPEED will be included in this study because it is a registry data source that provides 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 will be 938.7k.

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

There are 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 will be 01 January 2015 (01 January 2016 with 365 days of prior data availability prior data availability prior to cohort entry data requirement) to study period end date.
Lastly, ethics approval for HI-SPEED is estimated to take 30 days from ethics submission of protocol, which makes the execution of this study feasible within the current study timelines.

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

SIDIAP will be included in this study because it is a primary care and hospital data source that provides 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 will be 794k.

Moreover, data availability and follow-up in SIDIAP is sufficient, as data availability starts in 2006, and the date of most recent data extraction is 30 June 2023, which aligns with the study period. The median follow-up of the first observation period in SIDIAP is 5,670 days.

There are no study specific limitations present in SIDIAP.
Lastly, IRB approval for SIDIAP is estimated to take 2 months, which makes the execution of this study feasible within the current study timelines.

Clinical Practice Research Datalink GOLD (CPRD GOLD)

CPRD GOLD will be included in this study because it is a primary care data source that provides 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 will be 811.2k.

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

There are 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 has blanket approval which makes the execution of this study feasible within the current 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 enables the use of standardised analytics and using DARWIN EU® tools across the network, since the structure of the data and the terminology system is harmonised. The OMOP CDM was developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel and in The Book of OHDSI: http://book.ohdsi.org.

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

Data storage and protection

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

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

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

QUALITY CONTROL

Data source quality control

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

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

PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

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

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

ANNEX IV. Operational definition of AESIs

Table S1. Preliminary codelist of Guillain-Barré syndrome definitions.

Concept ID	Concept name	Domain	Vocabulary
4164770	Guillain-BarrÃƒÂ© 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. Preliminary 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. Preliminary 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

4072086	Acute bloody pericarditis	Condition	SNOMED
4070323	Acute effusive pericarditis	Condition	SNOMED
315293	Acute idiopathic pericarditis	Condition	SNOMED
4304392	Acute mediastinopericarditis	Condition	SNOMED
4006483	Acute myopericarditis	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. Preliminary 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
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
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. Preliminary 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
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

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. Preliminary 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

Concept ID	Concept name	Domain	Vocabulary
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
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

Concept ID	Concept name	Domain	Vocabulary
380322	Postvaricella encephalitis	Condition	SNOMED
380632	Meningococcal encephalitis	Condition	SNOMED
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
4047623	Chronic viral encephalitis	Condition	SNOMED
4047625	Fungal encephalitis	Condition	SNOMED
4047626	Focal encephalitis	Condition	SNOMED

Concept ID	Concept name	Domain	Vocabulary
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
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

Concept ID	Concept name	Domain	Vocabulary
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
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

Concept ID	Concept name	Domain	Vocabulary
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
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. Preliminary 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
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

Concept ID	Concept name	Domain	Vocabulary
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
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

Concept ID	Concept name	Domain	Vocabulary
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
138723	Acquired red cell aplasia	Condition	SNOMED
137829	Aplastic anemia	Condition	SNOMED
137829	Aplastic anemia	Condition	SNOMED
138723	Acquired red cell aplasia	Condition	SNOMED
140681	Constitutional aplastic anemia	Condition	SNOMED
318397	Chronic idiopathic thrombocytopenic purpura	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
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

Concept ID	Concept name	Domain	Vocabulary
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
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

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

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

ANNEX V. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

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

EU PAS Register® number: EUPAS1000000969
Study reference number (if applicable): P4-C2-018

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☒	☐	☐	
1.1.2 End of data collection[2]	☒	☐	☐	
1.1.3 Progress report(s)	☐	☐	☒	
1.1.4 Interim report(s)	☐	☐	☒	
1.1.5 Registration in the EU PAS Register®	☐	☐	☒	
1.1.6 Final report of study results.	☒	☐	☐	

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☐	☐	☐	
2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	
2.1.2 The objective(s) of the study?	☒	☐	☐	
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

[1] Date from which information on the first study is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.
[2] Date from which the analytical dataset is completely available.

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	8.1
3.2 Does the protocol specify whether the study is based on primary, secondary, or combined data collection?	☒	☐	☐	8.2
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	8.8
3.4 Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☒	
3.5 Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☒	☐	☐	12

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	☒	☐	☐	8.5
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	☒	☐	☐	8.3
4.2.2 Age and sex	☒	☐	☐	8.5
4.2.3 Country of origin	☒	☐	☐	8.2
4.2.4 Disease/indication	☒	☐	☐	8.6
4.2.5 Duration of follow-up	☒	☐	☐	8.4
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	8.5

Comments:

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.1 Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☐	☐	☒	
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☐	☒	
5.3 Is exposure categorised according to time windows?	☐	☐	☒	

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

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	8.6.2
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.6.2
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☐	☒	
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

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

Comments:

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

Comments:

Section 9: Data sources	Yes	No	N/A	Section Number
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☐	☐	☒	
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	8.2
9.1.3 Covariates and other characteristics?	☒	☐	☐	8.2
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☐	☐	☒	
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	8.2
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	8.2
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☐	☐	☒	
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	8.2
9.3.3 Covariates and other characteristics?	☒	☐	☐	8.2
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☒	

Comments:

Section 10: Analysis plan	Yes	No	N/A	Section Number
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.8
10.2 Is study size and/or statistical precision estimated?	☐	☐	☒	
10.3 Are descriptive analyses included?	☒	☐	☐	8.8

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.4 Are stratified analyses included?	☒	☐	☐	8.8
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	
10.7 Does the plan describe methods for handling missing data?	☐	☐	☒	
10.8 Are relevant sensitivity analyses described?	☒	☐	☐	8.8

Comments:

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

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	☐	☒	☐	
12.1.2 Information bias?	☒	☐	☐	9
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☐	☒	☐	
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	8.7

Comments:

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

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

Comments:

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

Comments:

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

Comments:

Name of the main author of the protocol: Ivan Lam

Date: 21/01/2026

Signature: Ivan Lam

ANNEX VI. Glossary

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

Aggregated Data

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

Benefit-Risk Assessment

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

Common Data Model (CDM)

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

Complex Studies (C3)

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

Coordination Centre (CC)

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

Data Access

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

Data Quality Framework

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

Data Source

A data source 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 data sources 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 data sources, 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.