Study Protocol

P3-C3-001

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

27/11/2024

Version 4.0

Table of Contents

DOCUMENT HISTORY

VERSION	DATE	DESCRIPTION
V1.0	06/06/2024	**Draft 1 Protocol version for EMA review**
V2.0	04/07/2024	**Draft 2 Protocol submitted for EMA review**
V3.0	10/07/2024	**Archiving version of the protocol submitted to EMA**
V4.0	27/11/2024	**Amendment to add a new study team member (Section 2)**

Study Title	DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe
Protocol version identifier	V4.0
Date of last version of protocol	27/11/2024
EU PAS register number	EUPAS1000000254
Active substance	n/a
Medicinal product	n/a
Research question and objectives	This study aims to estimate the background incidence rates of selected vaccine adverse events of special interest, as well as to understand the demographic and clinical characteristics of patients in Europe. 1. To estimate population level incidence rates of selected adverse events of special interest (AESIs) in the general population during 2010 and until the latest data availability, stratified by calendar year, month, sex, age groups, and data source. 2. To describe demographic and clinical characteristics of individuals with incident AESIs, and to compare the characteristics with individuals of similar age and sex but without the AESI.
Country(ies) of study	UK, the Netherlands, Spain, Norway, Germany
Author	Xintong Li, Edward, Burn, Albert Prats-Uribe

LIST OF ABBREVIATIONS

Abbreviation	Name
AESI	Adverse events of special interest
CDM	Common Data Model
COVID-19	Coronavirus disease-2019
EHR	Electronic Health Record
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
OMOP	Observational Medical Outcomes Partnership
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
RT-PCR	Reverse transcription polymerase chain reaction
CPRD	Clinical Practice Research Datalink
IPCI	Integrated Primary Care Information
SIDIAP	The Information System for Research in Primary Care
SNOMED	Systematized Nomenclature of Medicine
SPEAC	Safety Platform for Emergency vACcines

1. TITLE

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

2. RESPONSIBLE PARTIES – STUDY TEAM

STUDY TEAM ROLE	NAMES	ORGANISATION
Study Project Manager/Principal Investigator	Xintong Li, Edward Burn	University of Oxford
Data Scientist	Edward Burn	University of Oxford
Epidemiologist	Xintong Li Amy Lam	University of Oxford University of Oxford
Clinical Domain Expert	Daniel Prieto-Alhambra Albert Prats-Uribe George Corby James Bezer Abigail Robinson Ffion Samuels	University of Oxford University of Oxford University of Oxford University of Oxford University of Oxford University of Oxford
Data Partner*	**Names**	**Organisation**
	Antonella Delmestri Mees Mosseveld Talita Duarte Salles James Brash Hedvig Nordeng	University of Oxford – CPRD data Erasmus MC – IPCI data IDIAPJGol – SIDIAP data IQVIA – Germany NLHR Norway – Uni Oslo

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

3. ABSTRACT (STAND-ALONE SUMMARY OF THE STUDY PROTOCOL)

Title

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

Rationale and background

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

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

Research question and objectives

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

Main objectives

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

Secondary objective

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

Study design

Population-level cohort study.

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 2010 or the date that individual fulfil the data availability and outcome 'clean window' requirement.

Variable:

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

Data source:

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

Analysis:

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

For objective 1 incidence rates per 100,000 person-years and corresponding 95% confidence intervals will be reported using exact Poisson model. Analysis will be stratified by calendar month, 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 for Objective 2.

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

4. AMENDMENTS AND UPDATES

Number	Date	Section of study protocol	Amendment or update	Reason
V3.1	26/11/2024	2.RESPONSIBLE PARTIES – STUDY TEAM	Amendment	Adding one epidemiologist to the study team

5. MILESTONES

STUDY SPECIFIC MILESTONES	TIMELINE
Draft Study Protocol	May 2024
Final Study Protocol	June 2024
Creation of Analytical code/ Phenotype evaluation/ IRB Approvals	August / September 2024
Execution of Analytical Code on the data	October 2024
Draft Study Report	November 2024
Final Study Report	November / December 2024

6. RATIONALE AND BACKGROUND

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

A rapid initial regulatory response to a vaccine safety concern is crucial for maintaining public confidence in vaccination programs. Background incidence rates of adverse events of special interest (AESIs) may support the rapid initial response to a vaccine safety concern.(Black et al. 2021) Observed-to-expected analyses are essential for such a response, in order to inform further steps of signal strengthening and evaluation.(van der Boom, van Eekeren, and van Hunsel 2023; Mahaux, Bauchau, and Van Holle 2016; Gordillo-Marañón et al. 2024).

In 2020, the EMA-funded Vaccine Covid-19 monitoring readiness (ACCESS, EUPAS37273) project aimed at estimating the background rates AESIs for monitoring COVID-19 vaccines.(Willame et al. 2021; "Background Rates of Adverse Events of Special Interest for Monitoring COVID-19 Vaccines (ACCESS-BGR)" n.d.) Several publications from other research groups have also contributed to the global knowledge regarding background incidence rates of AESIs.(Li et al. 2021; Willame et al. 2023) However, there is a need for regular updates to support readiness in case of a new safety concern, regardless of the vaccine product. There is also a need for higher granularity of estimates, especially with regards to groups at risk (e.g., age, comorbidities), as well as other factors such as seasonality/circulation of specific virus strains considered as independent risk factors for some events (e.g., addressed by stratification by month).

Previous studies have shown that background rates vary across age groups and sex, and are often heterogeneous between regions and data sources.(Li et al. 2021; Ostropolets et al. 2022; Willame et al. 2023)

The heterogeneity can come from different clinical coding systems, health care delivery system, clinical practice, or reflect the true differences between the source population. It is therefore important to use the background rates that generated from the same or similar data source rather than rates estimated from different setting or data sources in observed-to-expected analysis. Understanding the demographic and clinical characteristics of patients could provide useful information for evaluating potential safety signals in the further.(Burn et al. 2022)

While some AESIs are considered specific to given vaccines, vaccine platforms, or classes of vaccines, several, such as Guillain-Barre syndrome (GBS) and other immune-mediated or neurological outcomes are listed across a wide range of vaccines. Therefore, the current study aims to expand the scope of previous studies to not only AESIs for approved vaccines, evidence generated from this study can also be used to support further newly developed vaccines. This study will support vaccine safety monitoring endeavours as part of the Vaccine Monitoring platform("Vaccine Monitoring Platform" 2023).

7. RESEARCH QUESTION AND OBJECTIVES

Table 1. Primary and secondary research questions and objective.

A. Primary research questions and objectives.

Objective:	To estimate population level incidence rates of selected vaccine adverse events of special interest (AESIs) in the general population from 2010 and until the latest data availability, stratified by calendar year, month, sex, age groups, and data source
Hypothesis:	n/a
Population *(mention key inclusion-exclusion criteria):*	The general 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 2010 or the date that individual fulfil the data availability and the 'clean window' requirement.
Exposure:	n/a
Comparator:	n/a
Outcome:	Incident cases of pre-specified AESIs
Time *(when follow up begins and ends):*	2010 to the latest data availability of each participated database
Setting:	Primary care electronic health records from CPRD GOLD [UK], IPCI [The Netherlands], IQVIA DA Germany [Germany], NLHR [Norway], and SIDIAP [Spain].
Main measure of effect:	Incidence rates

Objective:	To estimate age and sex standardised (to the European population) population level incidence rates of selected vaccine adverse events of special interest (AESIs) in the general population from 2010 and

	until the latest data availability, stratified by calendar year, month, sex, age groups, and data source
Hypothesis:	Same as above
Population *(mention key inclusion-exclusion criteria):*	Same as above
Exposure:	Same as above
Comparator:	Same as above
Outcome:	Same as above
Time *(when follow up begins and ends):*	Same as above
Setting:	Same as above
Main measure of effect:	Age and sex standardised incidence rates

B. Secondary research question and objective.

Objective:	To describe demographic and clinical characteristics of individuals with incident AESIs, and comparing the characteristics with individuals of similar age and sex but without the AESI.
Hypothesis:	n/a
Population *(mention key inclusion-exclusion criteria):*	Individuals with observed AESIs: Cohorts of individuals with each study event of interested will be constructed. The index date will be the date of diagnosis or confirmatory test for the disease diagnosis. Matched cohort of individuals without AESI: For each AESI cohort, we will construct an age and sex matched cohort from the general population without that AESI. We will require individuals in the matched cohort to be under observation on the index date of the AESI cohort. This matched cohort will be used to compare the patient characteristics with the AESI cohort.
Exposure:	Individuals with incident pre-specified AESIs
Comparator:	Individuals without the pre-specified AESI
Outcome:	n/a
Time *(when follow up begins and ends):*	Index date will be the date of the incident AESIs. Characteristics will be summarised for time prior to index, no follow up time will be involved.
Setting:	Primary care electronic health records from CPRD GOLD [UK], IPCI [The Netherlands], IQVIA DA Germany [Germany], NLHR [Norway], and SIDIAP [Spain].
Main measure of effect:	n/a

8. RESEARCH METHODS

8.1 Study type and Study Design

Table 2. Description of Potential Study Types and Related Study Designs.

STUDY TYPE	STUDY DESIGN	STUDY CLASSIFICATION
Population-level descriptive epidemiology Patient-level characterisation	Population-level cohort	Complex (C3) *

*Note: This study is classified as C3: Complex because of the high volume and the complexity of the phenotypes needed to be generated for the study.

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

8.2 Study Setting and Data Sources

Database selection

The selection of databases for this study was performed based on data reliability and relevance for the proposed research question among those databases onboarded and available within DARWIN EU. The selected databases fulfil the criteria required for the availability of key information on exposures (i.e. complete recording of vaccines including date and brand), outcomes (except for hospitalization), and covariates, while covering different settings and regions of Europe.

Potential for future study repetition

This study could be replicated in the future to include additional databases provided those are onboarded for DARWIN EU and fulfil the selection criteria as outlined above.

Information on data source(s) planned to be used with a justification for their choice in terms of ability to capture the relevant data is described in **Table 3**.

Table 3. Description of the selected Data Sources.

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active subjects*	Feasibility count of disease (if relevant)**	Data lock for the last update
UK	Clinical Practice Research Datalink (CPRD) GOLD	Data with denominator populations representative of the general source population provide the most relevant and valid background rates.	Primary care	EHR	17.4 million	CVST: 1300 deep venous thrombosis: 183700 GBS: 3500 Haemorrhagic stroke: 23100 Immune thrombocytopenia: 11200 Ischemic stroke: 48400 Acute myocardial infarction: 295000 narcolepsy: 3600 Pulmonary embolism: 114100 thrombocytopenia: 52600 Arrhythmia: 38800 Rheumatoid arthritis: 218700 Type 1 diabetes mellitus: 106800 Tinnitus: 379600 Encephalitis: 900 Immune thrombocytopenia: 10700 Pericarditis: 100	01-01-2024
The Netherlands	Integrated Primary Care Information (IPCI)	Data with denominator populations representative of the general source population provide the most relevant and valid background rates.	Primary care	EHR	2.82 million	deep venous thrombosis: 61500 GBS: 2400 Haemorrhagic stroke: 14600 Ischemic stroke: 95300 Acute myocardial infarction: 364300 Pulmonary embolism: 74200 thrombocytopenia: 6100 Arrhythmia: 11800 Rheumatoid arthritis: 127800	30-04-2024

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active subjects*	Feasibility count of disease (if relevant)**	Data lock for the last update
						Type 1 diabetes mellitus: 86000 Tinnitus: 122900 Encephalitis: 6300 Pericarditis: 5700	
Spain	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP)	Data with denominator populations representative of the general source population provide the most relevant and valid background rates.	Primary care database + linkage to hospital data	EHR	8.55 million	CVST: 700 deep venous thrombosis: 106100 GBS: 11600 Haemorrhagic stroke: 101500 Immune thrombocytopenia: 20200 Ischemic stroke: 448400 Acute myocardial infarction: 420700 narcolepsy: 2400 Pulmonary embolism: 120400 thrombocytopenia: 202200 Arrhythmia: 56200 Rheumatoid arthritis: 111400 Type 1 diabetes mellitus: 200 Tinnitus: 238600 Immune thrombocytopenia: 20200 Pericarditis: 1500 Autoimmune thyroiditis: 25500	2023-06-30
Norway	Norwegian Linked Health Registry data (NLHR)	Linked national-wide population-level health registries.	Primary care, including specialists, and second	registries	5.67 million	deep venous thrombosis: 800 GBS: 100 Haemorrhagic stroke: 26100 Immune thrombocytopenia: 100 Ischemic stroke: 279300 Acute myocardial infarction: 540900 narcolepsy: 300	2021-12-31

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active subjects*	Feasibility count of disease (if relevant)**	Data lock for the last update
			ary care (hospital data)			Pulmonary embolism: 267300 thrombocytopenia: 700 Arrhythmia: 21600 Rheumatoid arthritis: 1100800 Type 1 diabetes mellitus: 1431400 Tinnitus: 156700 Immune thrombocytopenia: 100 Pericarditis: 900 Autoimmune thyroiditis: 300	
Germany	IQVIA DA Germany	Covers primary and secondary care setting with denominator populations representative of the general source population	Primary care (GP and specialist)	EHR	41.9 million	deep venous thrombosis: 209100 GBS: 13300 Haemorrhagic stroke: 106100 Immune thrombocytopenia: 11700 Ischemic stroke: 528100 Acute myocardial infarction: 618000 narcolepsy: 20800 Pulmonary embolism: 272300 thrombocytopenia: 176700 Arrhythmia: 92200 Rheumatoid arthritis: 1695100 Type 1 diabetes mellitus: Tinnitus: Encephalitis: Immune thrombocytopenia: 11700 Pericarditis: 8500	30-09-2023

* Number of Patients as reported in DARWIN EU Portal.

** Feasibility counts were estimated from record counts of concepts from the DARWIN Portal.

Clinical Practice Research Datalink GOLD, United Kingdom (University of Oxford)

The Clinical Practice Research Datalink (CPRD) GOLD is a database of anonymised electronic health records (EHR) from General Practitioner (GP) clinics in the UK that use the Vision® software system for their management.1 The source population encompasses 98% of the UK, registered with GPs responsible for non-emergency care and referrals. Participating GPs provide CPRD EHR for all registered patients who did not specifically request to opt out of data sharing. Covering 4.6% of the current UK population, GOLD includes 4.9% of contributing GP practices, providing comprehensive information within its defined source population. GOLD contains data from all four UK constituent countries and the current regional distribution of its GP practices is 5.7% in England, 55.6% in Scotland, 28.4% in Wales, and 10.2% in Northern Ireland (May 2022). GOLD data include patient's demographic, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.

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

The Integrated Primary Care Information (IPCI) database is a longitudinal observational database containing routinely collected data extracted from computer-based patient records of a selected group of general practitioners (GPs) across the Netherlands.1 IPCI was started in 1992 by the department of Medical Informatics of the Erasmus University Medical Center in Rotterdam. The current database includes patient records from 2006 on, when the size of the database started to increase significantly. The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex. Although the geographical spread is limited, GP practices are located in urban and non-urban areas.

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

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

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

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

Norwegian Linked Health Registry data (NLHR), Norway

Norway has a universal public health care system consisting of primary and specialist health care services covering a population of approximately 5.4 million inhabitants. Many population-based health registries were established in the 1960s with use of unique personal identifiers facilitating linkage between registries. Data in these health registries are used for health analysis, health statistics, improving the quality of

healthcare, research, administration and emergency preparedness. We harmonized data from the following registries: the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Surveillance System for Communicable Diseases (MSIS), the Norwegian Immunisation Registry (SYSVAK), the National Death Registry, and the National Registry (NR). Linkage between the registries was facilitated using project-specific person ID generated from unique personal identification assigned at birth or immigration for all legal residents in Norway.

IQVIA Disease Analyser (DA) Germany, Germany (IQVIA)

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

The database contains demographics records, basic medical data, disease diagnosis according to International Classification of Diseases, 10th revision (ICD-10), and prescription records.2 While the database partly records information on deaths and procedures, it currently does not support linkage with external data sources. Routine updates are conducted at regular intervals. IQVIA DA Germany is suitable for pharmacoepidemiologic and pharmacoeconomic studies as previously demonstrated.2,5,6 The quality of data is assessed based on several criteria including completeness of information and correctness (e.g. linkage between diagnosis and prescriptions).

8.3 Study Period

From January 2010– until last available data, depending on the data sources.

8.4 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 2010),

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

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

Individuals 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.

For acute and recurrent events. 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 4. 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 2010), 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	event specific (Table X)	IP, OP	No record of the same condition in the wash out period.
Sample of individuals from the general population matched to the individuals with observed AESIs	Date of diagnosis of the matched AESI individuals	multi	General population	No washout	IP, OP	-

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

8.5 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 2010 or the date that individual fulfil the data availability and outcome 'clean window' (details in 8.6.2 Outcomes) requirement.

8.6 Variables

8.6.1 Exposure/s

Not applicable.

8.6.2 Outcomes

AESIs:

The list was built on previously internationally recognized lists of AESIs as Brighton Collaboration/ Safety Platform for Emergency vACcines (SPEAC) and curated by experts from EMA and EU committees. Apart from AESIs that have been included in previous studies, a broader list of conditions has been added. For example, conditions related to skin reactions are included. We excluded AESIs specific to one vaccine only and already well characterised (e.g., intussusception) or those which are very rare (e.g., MIS).

The selected outcomes of interest are listed below. The process of generating the phenotypes is detailed in 8.8 Analysis.

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 5** for acute and recurrent outcomes. For chronic outcomes we will consider the event being incident if there is no record of that outcome anytime prior in the patient history. Also, the time will be censored after the first occurrence of a chronic event and the patient will not be allowed to re-enter the cohort.

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.

The event-specific clean window is subject to changes during the phenotyping stage and will therefore be specified in the Methods section of the results report. After diagnosing the created cohorts, the timeframes might change, and these adjustments will be detailed in the report.

Table 5. Summary of event specific clean windows for outcomes

	Outcome	Event specific clean window**	Type
Immune-mediated diseases	Guillain Barré syndrome	90	Acute
	Kawasaki disease	90	Acute
	Narcolepsy	NA	Chronic
	Immune Thrombocytopenia	90	Acute
	Type 1 diabetes	NA	Chronic
	Autoimmune thyroiditis	NA	Chronic
	Facial nerve palsy/Bells' palsy	90	Acute
Blood disorders	Thrombocytopenia	90	Acute
	Thrombotic thrombocytopenia syndrome	90	Acute
Cardiovascular system	Coronary artery disease	NA	Chronic
	Heart failure	NA	Chronic
	Single organ cutaneous vasculitis	NA	Chronic
	Arrhythmia	NA	Chronic
	Thrombotic microangiopathy $$	90	Acute
	Cardiomyopathy	NA	Chronic
	Myocarditis	90	Acute

	Outcome	Event specific clean window**	Type
	Pericarditis	90	Acute
	Myocarditis/Pericarditis	90	Acute
	Disseminated intravascular coagulation	90	Acute
	Deep vein thrombosis	90	Acute
	Pulmonary embolism	90	Acute
	Ischaemic stroke	90	Acute
	Haemorrhagic stroke	90	Acute
	Cerebral venous thrombosis	90	Acute
	Epileptic convulsions/seizures	90	Acute
Nervous system	Non-epileptic convulsions/seizures	90	Acute
	Febrile seizure	30	Recurrent
	Multiple sclerosis	NA	Chronic
	Acute Aseptic Meningitis	90	Acute
	Myelitis including transverse myelitis	90	Acute
	Encephalitis, which includes encephalomyelitis and ADEM	90	Acute
	Neuritis including optic neuritis	90	Acute
	Erythema multiforme	90	Acute
Skin, bones and joints system	Rheumatoid arthritis	NA	Chronic
	Drug reaction with eosinophilia and systemic symptoms	90	Acute
	Stevens-Johnson syndrome/ Toxic epidermal necrolysis	90	Acute
	Acute kidney injury	90	Acute
Others	Acute liver injury	90	Acute
	Anaphylaxis	30	Recurrent
	Pancreatitis (Acute)	90	Acute
	Rhabdomyolysis	90	Acute
	Sensorineural hearing loss	NA	Chronic
	Tinnitus	30	Recurrent
	Postmenopausal haemorrhage	30	Recurrent

**Event specific clean window: For each study outcome, a specific clean window will be applied to define incident outcomes. This definition is subject to changes during the phenotype stage. NA identifies those outcomes that are chronic, and we will consider that the washout window is lifetime.

$ Outcome types include: acute, chronic, recurrent, etc. Further details will be provided during the phenotype stage.

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

Demographics:

Sex: Female, male.

Age

Age groups:

0 - 1
2 - 4
5 - 11
12 - 17
18 - 29
30 - 39
40 - 49
50 - 59
60 - 64
65 - 79
80 and over

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

Health conditions:

We will use large-scale characterisation to identify individuals´ history of the comorbidities using all available data prior to the index date (date of entering the general population cohort, or date of incident event). This is a data-driven method where all available data in the dataset will be utilized to provide a comprehensive view of the potential comorbidities of an individual. Details of the large-scale characteristics is explained in Section 8.8 Analysis.

Medication use:

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

The operational definition of the covariates is described in **Table 6 Table** .

Table 6. Operational Definitions of Covariates.

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type	Diagnosis Position[2]	Applied to study populations	Measurement characteristics/ validation	Source for algorithm
Comorbidity	Large-scale patient-level characterisation with regard to underlying comorbidities	Binary	[Inf, -1] [-365, -1]	IP, OP, OT	SNOMED	n/a	All	n/a	n/a
Medication pre-index	Large-scale patient-level characterisation regarding use of medications	Binary	[-183, -1]	IP, OP, IT	ATC, RxNorm	n/a	all	n/a	n/a

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

8.7 Study size

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

8.8 Analysis

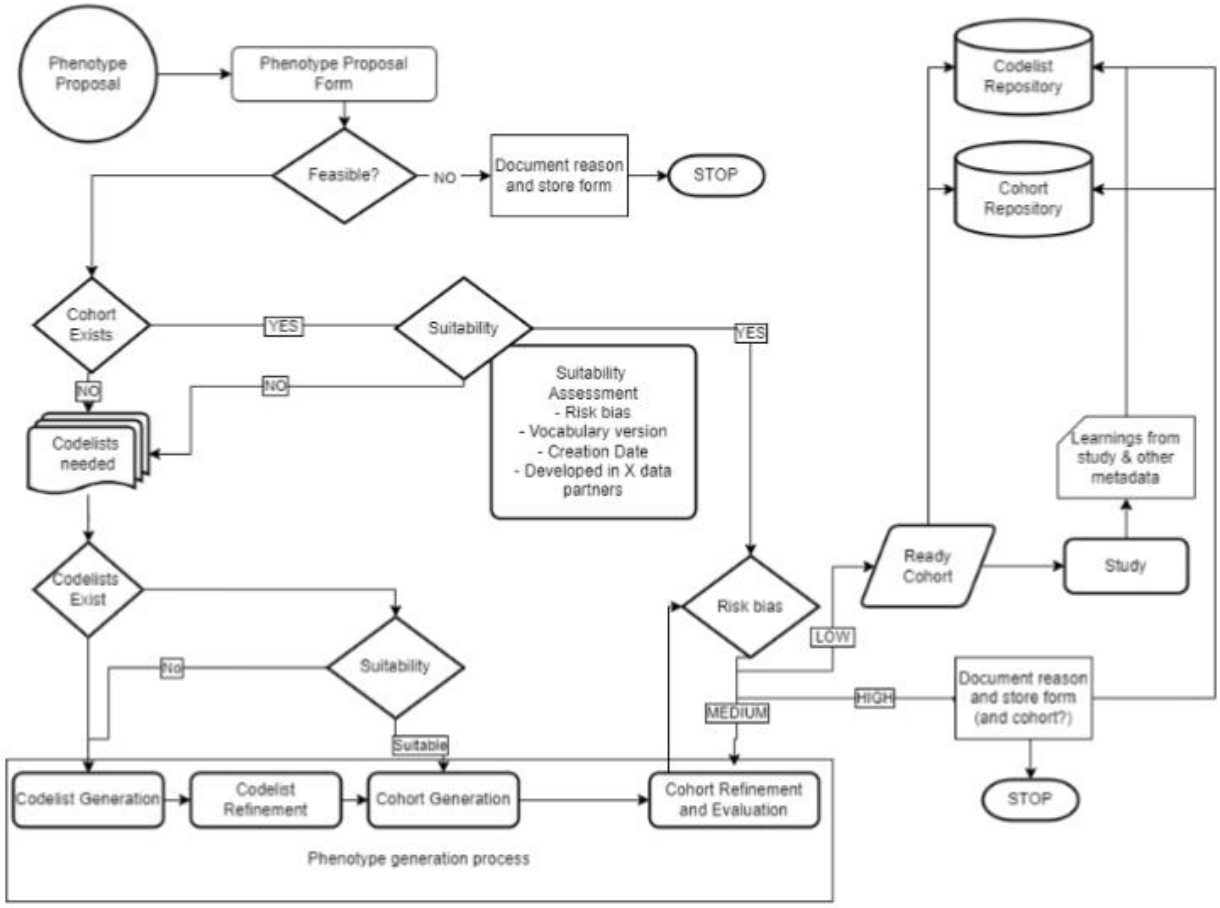
The analyses will be conducted in two stages. First, we will develop the phenotypes for the AESIs in this study. Then, we will estimate the incidence rates of the AESIs.

Stage 1: Phenotyping the AESIs

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

According to Hripcsak and Albers (Hripcsak and Albers 2018)"a phenotype is a specification of an observable, potentially changing state of an organism, […]. The term phenotype can be applied to patient characteristics inferred from electronic health record (EHR) data. […] . Phenotype algorithms – i.e., algorithms that identify or characterize phenotypes – may be generated by domain exerts and knowledge engineers, including recent research in knowledge engineering or through diverse forms of machine learning […] to generate novel representations of the data."

Figure 1. Steps in the Phenotyping process in DARWIN EU studies



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

A detailed explanation of the sections present in *clinical description* is available in the appendix.

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

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

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

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

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

For each purposed phenotype, the following information will be documented:

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

Stage 2: Estimating the incidence rates of AESIs

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 will be applied to each AESI, an individual will contribute to the person-time after the clean window of the previous event been fulfilled, except if the event is chronic.

The incidence rates will be calculated using the "IncidencePrevalence" R package, developed by DARWIN EU.(Raventós et al. 2024)

Stratification/subgroup

- By database
- By calendar year
- By time Before / during/ after pandemic
- By age group (0 – 1, 2 – 4, 5 – 11, 12 – 17, 18 – 29, 30 – 39, 40 – 49, 50 – 59, 60 – 64, 65 – 79, 80 and over). 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.

These incidences will be further standardized by age and sex to the European population by the direct method, using the same age groups.

Objective 3: Population-level characteristics

For the general population cohort, we will use the large-scale characterisation to describe the characteristics on comorbidities and medication use, overall and for each calendar year.

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

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

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

Sensitivity analysis:

While in the main analysis where we will apply a 365-days clean window to all study events, in the sensitivity analysis, we will apply event specific clean window defined during the phenotyping process based on clinical plausibility.

Table 7. Primary, secondary, and subgroup analysis specification.

A. Primary analysis

Hypothesis:	Not applicable, descriptive incidence rate.
Exposure contrast:	n/a
Outcome:	AESIs as listed in section 8.6.2 Outcomes
Analytic software:	R
Model(s): ***(provide details or code)***	We will use Poisson models to 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	For incidence rates, we will estimate the rates with stratification.
Missing data methods	***Name method and provide relevant details, e.g. missing indicators, complete case, last value carried forward, multiple imputation (specify model/variables), other.***
	We will only include individuals with complete age and sex information.
Subgroup Analyses	***List all subgroups***
	Sex: male, female Age groups: 1) 0 – 1, 2 – 4, 5 – 11, 12 – 17, 18 – 29, 30 – 39, 40 – 49, 50 – 59, 60 – 64, 65 – 79, 80 and over. 2) 0-17, 18-64, 65+

Description of sensitivity analyses is presented in **Table 8**.

Table 8. 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
Clean window for outcome events	Event specific clean window, tailored to each event and clinically relevant will be defined during the phenotyping process, then applied in the sensitivity analysis	The definition of the clean window may impact on the included study population and outcome cases. If the event specific clean window is shorter, we might see higher incidence rates for that event.	A longer clean window may increase the specificity of incident event	Shorter clean window may include prevalent events

8.9 Evidence synthesis

With existing knowledge on the heterogeneity of background rates from different data sources, we will not synthesis the estimates across databases.

9. DATA MANAGEMENT

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

The analytic code for this study will be written in R and will use standardized analytics wherever possible. Each data partner will execute the study code against their database 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.

9.1 Data storage and protection

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

All databases used in this study are already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to. In agreement with these regulations, rather than combining person

level data and performing only a central analysis, local analyses will be run, which generate non-identifiable aggregate summary results.

The output files are stored in the DARWIN Remote Research Environment. These output files do not contain any data that allow identification of subjects included in the study. The RRE implements further security measures in order 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.

10. QUALITY CONTROL

General database quality control

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

Study specific quality control

Objective 1 of this study includes the phenotyping process, where we will create the phenotypes of the study outcomes. This will follow the standard procedure with all decision reviewed documented. We will reach out to the EMA for complex phenotypes.

11. LIMITATIONS OF THE RESEARCH METHODS

General limitations:

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

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

Misclassification of outcomes could happen if individuals received health care service outside of the data capture system. For example, in the UK primary care data, we would not be able to capture event that 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 diagnosis such as XXXX will have lower number of cases, if the diagnosis is mostly made in hospital.

We will report the data source-specific incidence rates.

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 databases, and further diagnostics would be needed when applying these phenotypes in other databases or later versions of the same databases.

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

12. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Adverse events/adverse reactions will not be collected or analysed as part of this evaluation. The nature of this non-interventional evaluation, through the use of secondary data, does not fulfil the criteria for reporting adverse events, according to module VI, VI.C.1.2.1.2 of the Good Pharmacovigilance Practices (https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf).

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

13. GOVERNANCE BOARD ASPECTS

SIDIAP, IPCI, and NLHR will require ethical approvals from their local Institutional Review Boards to perform this study. This study has been approved in CPRD under a previous protocol (23_003556: Population-level incidence, prevalence and mortality rates in diseases of regulatory interest).

14. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

Dissemination activities to be undertaken will include mainly, although not exclusively, the creation of a final report, scientific publications, and presentations at conferences.

15. OTHER ASPECTS

None.

16. REFERENCES

1. "Background Rates of Adverse Events of Special Interest for Monitoring COVID-19 Vaccines (ACCESS-BGR)." n.d. Accessed April 25, 2024. https://catalogues.ema.europa.eu/node/3578/administrative-details.
2. Black, Steven B., Barbara Law, Robert T. Chen, Cornelia L. Dekker, Miriam Sturkenboom, Wan-Ting Huang, Marc Gurwith, and Greg Poland. 2021. "The Critical Role of Background Rates of Possible Adverse Events in the Assessment of COVID-19 Vaccine Safety." Vaccine 39 (19): 2712–18.
3. Boom, Meiwen D. X. van der, Rike van Eekeren, and Florence P. A. M. van Hunsel. 2023. "Observed-over-Expected Analysis as Additional Method for Pharmacovigilance Signal Detection in Large-scaled Spontaneous Adverse Event Reporting." Pharmacoepidemiology and Drug Safety 32 (7): 783–94.
4. Burn, Edward, Xintong Li, Kristin Kostka, Henry Morgan Stewart, Christian Reich, Sarah Seager, Talita Duarte-Salles, et al. 2022. "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 31 (5): 495–510.
5. Gordillo-Marañón, María, Gianmario Candore, Karin Hedenmalm, Kate Browne, Robert Flynn, Loris Piccolo, Aniello Santoro, Cosimo Zaccaria, and Xavier Kurz. 2024. "Lessons Learned on Observed-to-Expected Analysis Using Spontaneous Reports during Mass Vaccination." Drug Safety: An International Journal of Medical Toxicology and Drug Experience, April. https://doi.org/10.1007/s40264-024-01422-8.
6. Hripcsak, George, and David J. Albers. 2018. "High-Fidelity Phenotyping: Richness and Freedom from Bias." Journal of the American Medical Informatics Association: JAMIA 25 (3): 289–94.
7. Li, Xintong, Anna Ostropolets, Rupa Makadia, Azza Shoaibi, Gowtham Rao, Anthony G. Sena, Eugenia Martinez-Hernandez, et al. 2021. "Characterising the Background Incidence Rates of Adverse Events of Special Interest for Covid-19 Vaccines in Eight Countries: Multinational Network Cohort Study." BMJ (Clinical Research Ed.) 373 (June): n1435.
8. Mahaux, Olivia, Vincent Bauchau, and Lionel Van Holle. 2016. "Pharmacoepidemiological Considerations in Observed-to-expected Analyses for Vaccines." Pharmacoepidemiology and Drug Safety 25 (2): 215–22.
9. Ostropolets, Anna, Xintong Li, Rupa Makadia, Gowtham Rao, Peter R. Rijnbeek, Talita Duarte-Salles, Anthony G. Sena, et al. 2022. "Factors Influencing Background Incidence Rate Calculation: Systematic Empirical Evaluation across an International Network of Observational Databases." Frontiers in Pharmacology 13 (April). https://doi.org/10.3389/fphar.2022.814198.
10. Raventós, Berta, Martí Català, Mike Du, Yuchen Guo, Adam Black, Ger Inberg, Xintong Li, et al. 2024. "IncidencePrevalence: An R Package to Calculate Population-level Incidence Rates and Prevalence Using the OMOP Common Data Model." Pharmacoepidemiology and Drug Safety 33 (1). https://doi.org/10.1002/pds.5717.
11. "Vaccine Monitoring Platform." 2023. European Centre for Disease Prevention and Control. March 8, 2023. https://www.ecdc.europa.eu/en/about-ecdc/what-we-do/partners-and-networks/eu-institutions-and-agencies/vaccine-monitoring.
12. Willame, C., C. Dodd, C. E. Durán, Rjhj Elbers, R. Gini, C. Bartolini, O. Paoletti, et al. 2023. "Background Rates of 41 Adverse Events of Special Interest for COVID-19 Vaccines in 10 European Healthcare Databases - an ACCESS Cohort Study." Vaccine 41 (1): 251–62.
13. Willame, C., C. Dodd, R. Gini, C. E. Durán, R. M. Thomsen, L. Wang, A. Gedebjerg, et al. 2021. "Background Rates of Adverse Events of Special Interest for Monitoring COVID-19 Vaccines." Zenodo. https://doi.org/10.5281/ZENODO.5255870.

17. ANNEXES

Appendix I: Clinical Description and code list for Kawasaki Disease

Appendix II: ENCePP checklist for study protocols

Appendix I:

Clinical description

Phenotype name: Kawasaki Disease (mucocutaneous lymph node syndrome)

Overview

Kawasaki disease is an acute, self-limited, systemic vasculitis of unknown aetiology, which predominantly (75% to 85% of cases) affects children aged <5 years. It appears more common in Japanese and black children, and in 25% of cases can cause coronary artery dilatations or aneurysms. In the UK it is estimated to affect around 8 children in every 100,000 under the age of 5.

Kawasaki disease is the second most common vasculitis in childhood (the most common being IgA vasculitis).

Background epidemiology

Kawasaki disease (KD) almost exclusively affects young children, with peak incidence between the ages of 13-24 months.[It is rare in the first 6 months of life, and 80% of all cases occur before age 5 years. It affects approximately 5000 US children annually with an incidence in the US of 24.7 per 100,000 children aged <5 years in 2010, based on hospitalisation data. The incidence in the UK is 4.5 per 100,000 children aged <5 years.[

Although KD has been described in all ethnicities throughout the world.[6] The incidence is highest in patients from Northeast Asia, especially Japan and Korea, suggesting a significant role for host genetics in pathogenesis. In the US, children of Asian/Pacific origin have the highest KD rate (50.4 per 100,000), followed by children of black (29.8) and white (22.5) origins. An earlier report from 1997 to 2000 showed an annual incidence of 16.9, 11.1, and 9.1 in 100,000 in the US for black, Hispanic, and white children aged <5 years, respectively.

Japanese KD surveillance studies have shown the incidence of KD is increasing relative to the decreasing birth rate.

Presentation

Initial signs include generally feeling unwell, fever, swollen lymph nodes, rash, redness / puffiness of palms and soles which later peel, bloodshot eyes, and redness on the inside of the lips and tongue and inside of the mouth. Some children can have painful joints, tummy ache, diarrhoea, and vomiting.

Around 20-40% of children with Kawasaki disease develop problems with heart including coronary aneurysms, valvular issues or pericardial effusions.

Assessment

Diagnostic criteria for Kawasaki disease, in children with ≥ 5 days of high fever:

Changes in lips and oral cavity (strawberry tongue, red cracked lips)

Bulbar conjunctivitis

Morbilliform rash

Hand and foot redness and swelling, or periungual peeling

Asymmetric anterior cervical adenopathy

Blood tests of use can include inflammatory markers (CRP, ESR), full blood count, electrolytes, liver transaminases, albumin and urinalysis.

Echocardiography and electrocardiography should be performed on all patients with Kawasaki disease or suspected/atypical Kawasaki disease to look for cardiac abnormalities

Management

Clinically, the course of untreated KD is divided into the following stages:

Acute febrile stage (lasting weeks 1-2)

Fever, irritability, cervical adenitis, conjunctivitis, rash, mucosal erythema, painful erythema of the hands and feet, arthralgia or arthritis, possible myocarditis, and pericarditis.

Subacute stage (lasting weeks 2-4)

Fever, rash, and lymphadenopathy have resolved; if fever persists there is an increased risk of cardiac complications; persistent irritability, poor appetite, and conjunctival injection; desquamation of extremities begins at this stage.

The patient may be completely asymptomatic if given intravenous immunoglobulin (IVIG). Periungual desquamation may be the only apparent clinical manifestation.

Cardiac abnormalities (coronary artery ectasia or aneurysms) may develop during this stage, and rarely, later in patients treated with IVIG.

Convalescent (lasting weeks 4-8)

All signs of inflammation have receded, and acute phase markers normalise.

If present, coronary artery ectasia or aneurysms may persist and enlarge.

Chronic stage (variable)

If present, coronary artery dilation may resolve.

However, coronary artery aneurysms may persist through adulthood. Such patients are at risk of subsequent coronary artery thrombosis, rupture, and myocardial infarction.

Complications and prognosis

The main complication is coronary artery aneurysm as described above.

The disease generally self-limits and resolves with treatment. The British Paediatric Surveillance Unit survey reports 19% of treated patients in the UK with KD still developed coronary artery aneurysms increasing to 39% in those aged under 1 year despite intravenous immunoglobulin (with even higher coronary artery aneurysm complication rates in those aged under 1 year. Some of these will resolve at 1-2 years (especially if small), whereas some will persist, and rarely will rupture. Overall mortality is likely <1%, and peaks between 15-45 days after fever onset.

Differentials (for exclusion)

Staphylococcal or streptococcal infection

Systemic juvenile idiopathic arthritis (systemic JIA)

Scarlet fever

Acute rheumatic fever

Toxic shock syndrome (TSS)

Staphylococcal scalded skin syndrome

Stevens-Johnson syndrome

Drug reaction

Rocky Mountain spotted fever

Measles

Multi-system inflammatory syndrome in children (MIS-C)

Keywords used for search:

"Kawasaki disease", "Kawasaki", "mucocutaneous lymph node"

Resulting code list and codes selected

Concept ID	Concept name	Domain	Included
618979	Myocarditis due to Kawasaki disease	Condition	**X**
35615119	Aneurysm of coronary artery due to and following acute febrile mucocutaneous lymph node syndrome	Condition	**X**
314381	Acute febrile mucocutaneous lymph node syndrome	Condition	**X**
4189803	Paracoccus kawasakiensis	Observation	
1340204	History of event	Observation	
314381	Acute febrile mucocutaneous lymph node syndrome	Condition	**X**

Appendix II: ENCePP checklist for study protocols

Doc.Ref. EMA/540136/2009

ENCePP Checklist for Study Protocols (Revision 4)

Adopted by the ENCePP Steering Group on 15/10/2018

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

EU PAS Register® number: n/a
Study reference number (if applicable):

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]	☒	☐	☐	8.3
1.1.2 End of data collection[2]	☒	☐	☐	8.3
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	☒	☐	☐	4

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)	☒	☐	☐	6
2.1.2 The objective(s) of the study?	☒	☐	☐	7
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	8.5
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)	☐	☐	☒	

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
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?	☐	☐	☒	
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
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?	☒	☐	☐	10
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?	☒	☐	☐	11
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?	☒	☐	☐	13
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)?	☒	☐	☐	14
15.2 Are plans described for disseminating study results externally, including publication?	☐	☒	☐	

Comments:

Name of the main author of the protocol: Xintong Li

Date: dd/Month/year 06/06/2024

Signature: Xintong Li