Study Protocol

P4-C1-014

DARWIN EU® - Neonatal seizures: Incidence, prevalence, patient characterisation, and treatments in European countries

05/12/2025

Version 2.0

Authors: Ellen Gerritsen, Dina Vojinovic

Public

CONTENTS

Study title	DARWIN EU® - Neonatal seizures: Incidence, prevalence, patient characterisation, and treatments in European countries
Protocol version	V2.0
Date	05/12/2025
EUPAS number	EUPAS1000000822
Active Substance	Anti-seizure medication: Phenobarbital, phenytoin/fosphenytoin, levetiracetam, midazolam, lidocaine, and carbamazepine, with additional drugs, such as pyridoxine, to be considered based on the results of cohort diagnostics.
Medicinal Product	Not applicable
Research question and objectives	Research question What is the incidence and prevalence of seizures in neonates across European countries, and how are the clinical presentation and treatment patterns characterised in this population? Study objectives 1. To estimate the incidence and prevalence of seizures in neonates. 2. To characterise the demographic and clinical profile of neonates at the time of seizure diagnosis including: • Sex • Age • Gestational age • Birth weight • Clinical manifestations of seizures • Diagnostic tools used (e.g., EEG) and, where feasible, availability of EEG modalities (such as conventional EEG, amplitude EEG, or video EEG) • Comorbidities and concurrent conditions, including those presenting a high risk for seizure development, such as hypoxic-ischemic encephalopathy, stroke or haemorrhage, infections, cortical malformations, errors of metabolism (acute and inborn), and genetic aetiologies • Comedication 3. To characterise i) treatments received for the primary diagnosis of seizures in neonates, ii) treatment patterns, and iii) duration of treatment. 4. To describe short-, mid-, and long-term outcomes among neonates diagnosed with seizures (including survival, related diagnosed conditions such as neurodevelopmental and neurological disorders) at 3 months, 6 months, 2 years, and 6–7 years of age. All analyses for objectives 1–4 will, where feasible, be stratified by gestational age (term, preterm). Analyses for objective 1 will also be stratified by sex.
Countries of study	France, Finland, Hungary, Sweden, United Kingdom
Authors	Ellen Gerritsen, e.gerritsen@darwin-eu.org Dina Vojinovic, d.vojinovic@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/terms	Description
ASM	Anti-seizure medication
ATC	Anatomical Therapeutic Chemical
CC	Coordinating centre
CDM	Common Data Model
CDW Bordeaux	Clinical Data Warehouse of Bordeaux University Hospital
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DOI	Declaration of Interests
DQD	Data Quality Dashboard
DRE	Digital Research Environment
DTZ	Data Transfer Zone
ED	Emergency Department
EEG	Electroencephalography
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
FinOMOP-HUS	Hospital District of Helsinki and Uusimaa
GDPR	General Data Protection Regulation
GP	General Practitioner
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage
ICD	International Classification of Diseases
IP	Inpatient
IQR	Interquartile range
IRB	Institutional Review Board
KM	Kaplan-Meier
N/A	Not applicable
NNRD	National Neonatal Research Database
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
PYs	Person-years
RxNorm	Medical prescription normalised
SNOMED	Systemised Nomenclature of Medicine
SUCD	Semmelweis University Clinical Data

Acronyms/terms	Description
WHO	World Health Organisation

1. TITLE

DARWIN EU® - Neonatal seizures: Incidence, prevalence, patient characterisation, and treatments in European countries.

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigators	Ellen Gerritsen Dina Vojinovic	IQVIA
Data Scientists	Akram Mendez Gargi Jadhav	IQVIA
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	**Names**	**Data Partner Organisation***
CDW Bordeaux	Guillaume Verdy	Centre Hospitalier Universite de Bordeaux
FinOMOP-HUS	Tiina Wahlfors Päivi Lesonen Eric Fey Kimmo Porkka Marianna Niemi	HUS - Helsinki University Hospital, Hospital District of Helsinki and Uusimaa
SUCD	Ágota Mészáros Héja Tibor Bagyura Zsolt István Kiss Loretta Zsuzsa	Semmelweis University
HI-SPEED	Huiqi Li Fredrik Nyberg	Swedish Medical Products Agency - Gothenburg University
NNRD	Muriel Ramalli Ricardo Ribas Neena Modi Kayleigh Ougham	Imperial College London

*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® - Neonatal seizures: Incidence, prevalence, patient characterisation, and treatments in European countries.

Rationale and background

Neonatal seizures are among the most common neurological emergencies in newborns and are typically symptomatic of acute central nervous system insults, such as hypoxic-ischaemic encephalopathy, stroke, or infection. Diagnosis is challenging due to the predominance of electrographic-only seizures, requiring electroencephalogram (EEG) confirmation. This study aims to generate real-world evidence on the occurrence, clinical presentation, and treatment of neonatal seizures across European data sources to support regulatory decision-making.

Research question and objectives

Research question

What is the incidence and prevalence of seizures in neonates across European countries, and how are the clinical presentation and treatment patterns characterised in this population?

Objectives

The specific objectives of this study are:

1. To estimate the incidence and prevalence of seizures in neonates.
2. To characterise the demographic and clinical profile of neonates at the time of seizure diagnosis including:
    - Sex
    - Age
    - Gestational age
    - Birth weight
    - Clinical manifestation of seizures
    - Diagnostic tools used (e.g., EEG) and, where feasible, availability of EEG modalities (such as conventional EEG, amplitude EEG, or video EEG)
    - Comorbidities and concurrent conditions, including those presenting a high risk for seizure development, such as hypoxic-ischaemic encephalopathy, stroke or haemorrhage, infections, cortical malformations, errors of metabolism (acute and inborn), and genetic aetiologies
    - Comedication
3. To characterise i) treatments received for the primary diagnosis of seizures in neonates, ii) treatment patterns, and iii) duration of treatment.
4. To describe short-, mid-, and long-term outcomes among neonates diagnosed with seizures (including survival and related diagnosed conditions such as neurodevelopmental and neurological disorders) at 3 months, 6 months, 2 years, and 6–7 years of age.

All analyses for objectives 1–4 will, where feasible, be stratified by gestational age (term, preterm). Analyses for objective 1 will also be stratified by sex.

Methods

Study design

This is a descriptive cohort study comprising three components:

- *Descriptive disease epidemiology:* A population-level cohort analysis will be conducted to estimate incidence and prevalence of seizures in neonates *(objective 1).*
- *Characterisation:* A patient-level cohort analysis will be used to characterise the demographic and clinical profile of neonates *(objective 2),* describe treatment received for the primary diagnosis and associated treatment patterns *(objective 3 (i, ii))*, and assess short-, mid-, and long-term outcomes among neonates diagnosed with seizures *(objective 4).*
- *Drug utilisation study*: A new drug user cohort design will be applied to estimate the duration of treatment among neonates receiving pharmacological therapy for seizures *(objective 3 (iii)).*

Population

Descriptive epidemiology study (objective 1): The study population will include all neonates, defined as infants from birth to 28 days of life, present in the data source between 1 January 2014 and 31 December 2024 (or the latest date available).

Characterisations (objectives 2, 3(i, ii) and 4: The study population will include all neonates with a first recorded diagnosis of seizures during the neonatal period, present in the data source between 1 January 2014 and 31 December 2024 (or the latest date available). For outcome analyses *(objective 4)*, only neonates diagnosed with seizures at least 90 days prior to the end of data availability in each data source will be included to ensure sufficient follow-up. Additionally, for *objective 4*, eligible neonates must have no prior recorded diagnosis of the outcome condition of interest before the index date (date of seizure diagnosis).

Drug utilisation study (objective 3(iii)): The study population will include neonates diagnosed with seizures who initiate anti-seizure pharmacological treatment during the neonatal period and who are present in the data source between 1 January 2014 and 31 December 2024 (or the latest date available). To ensure a new-user design, eligible neonates must have no prior record of anti-seizure medication use before the index date (date of treatment initiation).

Variables

Exposure:

Anti-seizure medication, including phenobarbital, phenytoin/fosphenytoin, levetiracetam, midazolam, lidocaine, and carbamazepine. Additional medication, such as pyridoxine, may be added based on the findings from cohort diagnostics and further exploration.

Outcomes:

Mortality, neurodevelopmental disorders, and neurological conditions, such as cerebral palsy, developmental delay, intellectual disability, behavioural disorders (including attention deficit hyperactivity disorder and autism spectrum disorder), neurosensory sequelae (including hearing loss and blindness), and epilepsy (where feasible type and/or syndromes).

Relevant covariates:

Comorbidities and concurrent conditions, including febrile convulsions, hypoxic-ischaemic encephalopathy, stroke or haemorrhage, infections, traumatic brain injury, cortical/cerebral malformations, errors of metabolism (acute and inborn), and genetic aetiologies.

Data sources

1. Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)
2. France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)
3. Hungary: Semmelweis University Clinical Data (SUCD)
4. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
5. United Kingdom: National Neonatal Research Database (NNRD)

Study size

No sample size has been calculated, as this is a descriptive study not designed to test a specific hypothesis. Based on a preliminary feasibility assessment, the expected number of person counts for neonatal seizures in the included data sources ranges from approximately 600 (CDW Bordeaux) to 3,300 (HI-SPEED).

Statistical analysis

Overall and annual incidence rates of seizures per 1,000 person-years, if appropriate (or 1,000 person-days), along with period prevalence of seizures *(objective 1),* will be estimated in the population of neonates. The choice of denominator will be based on the characteristics of the data source. Estimates will be presented overall and stratified by sex. Incidence rates will be reported together with 95% Poisson confidence intervals. The statistical analyses will be performed based on OMOP CDM mapped data using the R package *IncidencePrevalence*.

Characteristics of neonates will be described by means of large-scale characterisation and pre-specified patient-level characterisation *(objectives 2, 3(i))*. Categorical characteristics of interest will be reported as counts and proportions, while continuous variables will be reported as medians with interquartile ranges (IQR). Demographic characteristics, including age (days) at diagnosis and sex, will be assessed at the date of first recorded diagnosis of seizures in neonates (index date). Gestational age at birth and birth weight will be assessed. Large-scale characterisation, including characterisation of comorbidities and comedication, will be assessed anytime before to 1 day before the index date, as well as at the index date. Frequency of pre-specified clinical manifestations will be assessed at the index date and 7 days after the index date, while frequency of pre-specified conditions of interest, such as hypoxic-ischaemic encephalopathy, stroke or haemorrhage, infections, cortical malformations, errors of metabolism (acute and inborn), and genetic aetiologies, will be assessed anytime before to 1 day before the index date, at the index date, and 30 days after the index date. In addition, the availability of diagnostic tests (e.g., EEG) and, where feasible, EEG modalities (conventional EEG, amplitude-integrated EEG, or video EEG) will be evaluated anytime before to 1 day before the index date, at the index date, and 30 days after the index date. The number and percentage of neonates receiving anti-seizure medication and therapeutic hypothermia will be evaluated at the index date and 30 days after the index date. These analyses will be conducted using the *CohortCharacteristics* and *DrugUtilisation* R packages based on OMOP CDM mapped data.

The treatment pattern following the diagnosis of seizures *(objective 3(ii))* will be presented by Sunburst and Sankey diagrams, which will provide information on sequences of anti-seizure medication use during neonatal period. The statistical analysis will be performed based on OMOP CDM mapped data using the *TreatmentPatterns* R package.

Drug utilisation analyses *(objective 3 (iii))* will include all neonates diagnosed with seizures who initiated anti-seizure treatment and had no recorded use of medication of interest prior to the index date. Treatment duration will be estimated, and the minimum, p25, median, p75, and maximum will be provided. This analysis will be conducted using the *DrugUtilisation* R package based on OMOP CDM mapped data.

Overall survival will be characterised using Kaplan-Meier curves, and the occurrence of related outcomes, such as neurodevelopmental and neurological disorders (e.g., epilepsy, where feasible, type and/or syndromes of epilepsy, as well as other relevant conditions), will be assessed using cumulative incidence

(objective 4). Outcomes will be evaluated at predefined time points: 3 months, 6 months, 2 years, and 6–7 years of age. This analysis will be conducted using the *CohortSurvival* R package based on OMOP CDM mapped data.

All analyses for objectives 1–4 will, where feasible, be stratified by gestational age (term, preterm).

A minimum cell count of 5 will be used when reporting results, with any smaller count reported as “<5” and zero counts as “0”.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	December 2025
Creation of Analytical code	December 2025/January 2026
Execution of Analytical Code on the data	December 2025/January 2026
Draft Study Report	January 2026
Final Study Report	To be confirmed by EMA

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

6. RATIONALE AND BACKGROUND

Seizures are one of the most common neurologic emergencies in neonates.[1] Neonatal seizures occur during the neonatal period, defined from birth to the first 28 days of life.[2] For preterm neonates, this period of 28 days is added to the expected delivery date. Most neonatal seizures are symptomatic of an underlying acute central nervous system insult, such as hypoxic-ischaemic encephalopathy, intracranial haemorrhage, stroke, infection, or electrolyte disturbances.[1] Preterm and low birth weight neonates are more likely to suffer seizures than full-term neonates.[2]

Most neonatal seizures are electrographic-only events, although some are accompanied by clinical signs. Many of the visible signs of neonatal seizures, such as chewing motions or "bicycling" movements of the legs, also occur in healthy newborns.[2] For this reason, testing (using a (video-) electroencephalogram – EEG) is required to confirm the diagnosis.

As seizures in the neonatal period have been shown to have a focal onset, a division into focal and generalised is unnecessary. Seizures can have a motor (automatisms, clonic, myoclonic, tonic), non-motor (autonomic, behaviour arrest), or sequential presentation.[3, 4]

The neuronal injury caused by seizures in neonates often results in long-term neurodevelopmental sequelae, such as cerebral palsy, cognitive impairment, learning disabilities, and future epilepsy.[1] To control the seizures, anti-seizure medications (ASMs) may be prescribed.[2, 5] There are several options for ASMs. For instance, phenobarbitone, phenytoin/fosphenytoin, and levetiracetam are among the most commonly used in neonates, although usage patterns might differ by country. Drugs such as lidocaine and midazolam are used as infusions for seizures that are refractory (difficult to control).[5]

The incidence of seizures in neonates is challenging to estimate. Most population-based studies are based on clinical diagnosis of neonatal seizures, without EEG confirmation. The overall incidence has been reported as 1–4 per 1,000 live births, with much higher risk among preterm and low birth-weight neonates.[1]

To support and inform the ongoing work carried out by the neonatal working group of the Paediatric Committee and related paediatric investigation plan assessments, this study is expected to gather up-to-date evidence on the occurrence of neonatal seizures and characterise the clinical presentation of newborns diagnosed with seizures and treatments received using electronic healthcare data sources from European countries.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

What is the incidence and prevalence of seizures in neonates across European countries, and how are the clinical presentation and treatment patterns characterised in this population?

Research objectives

The specific objectives of this study are:

1. To estimate the incidence and prevalence of seizures in neonates.
2. To characterise the demographic and clinical profile of neonates at the time of seizure diagnosis including:
    - Sex
    - Age
    - Gestational age
    - Birth weight
    - Clinical manifestations of seizures
    - Diagnostic tools used (e.g., EEG) and, where feasible, availability of EEG modalities (such as conventional EEG, amplitude EEG, or video EEG)
    - Comorbidities and concurrent conditions, including those presenting a high risk for seizure development, such as hypoxic-ischaemic encephalopathy, stroke or haemorrhage, infections, cortical malformations, errors of metabolism (acute and inborn), and genetic aetiologies
    - Comedication
3. To characterise i) treatment received from the primary diagnosis of seizures in neonates, ii) treatment patterns, and iii) duration of treatment.
4. To describe short-, mid-, and long-term outcomes among neonates diagnosed with seizures (including survival and related diagnosed conditions such as neurodevelopmental and neurological disorders) at 3 months, 6 months, 2 years, and 6–7 years of age.

All analyses for objectives 1–4 will, where feasible, be stratified by gestational age (term, preterm). Analyses for objective 1 will also be stratified by sex.

8. RESEARCH METHODS

8.1. Study design

A cohort study will be conducted using routinely collected health data from 5 data sources from 5 countries across Europe and in 4 EU member states. The study will comprise the following parts:

- Descriptive disease epidemiology study will be conducted to address *objective 1* (**Figure 1a**).
- Characterisation will be conducted to address *objectives 2, 3(i, ii), and 4* (**Figure 1b**).
- Drug utilisation study will be conducted to address *objective 3(iii)* (**Figure 1c**).

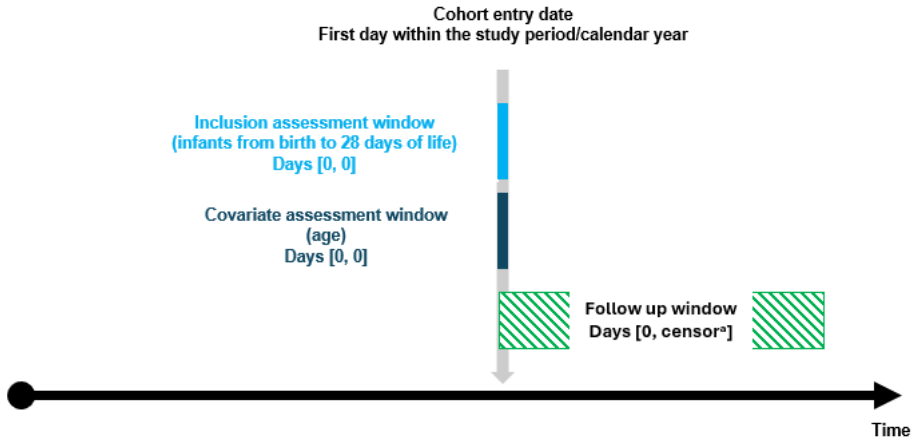


Figure 1a. Graphical depiction of the study design for objective 1.

a. Follow-up ends at the earliest of: end of neonatal period (28 days after birth), end of data source availability, end of the study period (31/12/2024), or death.

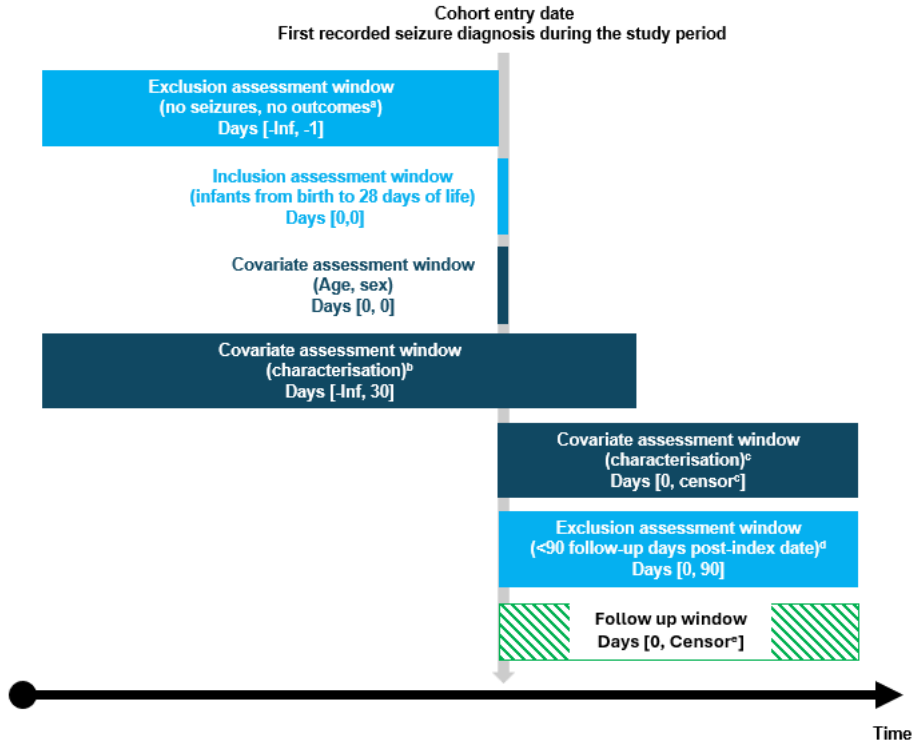


Figure 1b. Graphical depiction of the study design for objectives 2, 3(i, ii), and 4.

a. For outcome analyses (*objective 4*), eligible neonates must have no prior record of the condition of interest.
b. Comorbidities, comedication, pre-specified characteristics, and availability of diagnostic tools used.
c. Treatment received for the primary diagnosis and associated treatment patterns. For treatment pattern analysis , follow-up is censored at the end of the neonatal period (i.e., 28 days after birth), meaning that only pre-specified treatments administered within the neonatal period are included.
d. Exclusion criteria for short-, mid-, and long-term outcomes *(objective 4)*.
e. Death, end of data source availability, or end of the study period (31/12/2024).

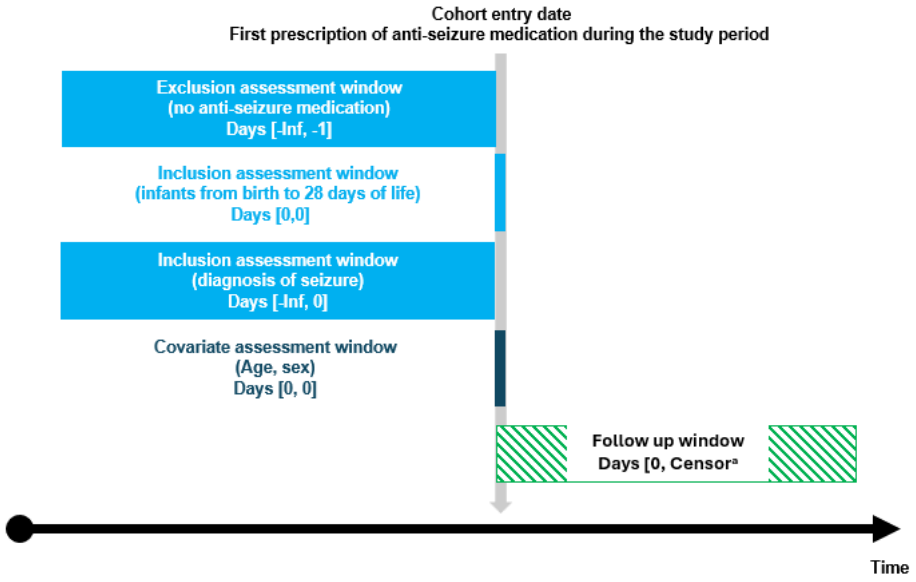


Figure 1c. Graphical depiction of the study design for objective 3(iii).

a. Death, end of data source availability, or end of the study period (31/12/2024).

8.2. Follow-up

For the descriptive disease epidemiology study assessing incidence and prevalence of seizures in neonates *(objective 1),* follow-up will start at the date of birth or the date when the individual meets all inclusion and exclusion criteria (index date). The study population will include infants present in the data source between 1 January 2014 and 31 December 2024 (or the latest available date in the data source). Follow-up will end at the earliest of: end of the neonatal period (28 days after birth), death, loss to follow-up, end of data source availability, or end of study period. For hospital-based data sources, follow-up will end at discharge, the end of the neonatal period, or the end of observation period as defined by the latest recorded event in the data source during the neonatal period, whichever occurs first. For incidence analysis, follow-up will also end upon the first recorded diagnosis of neonatal seizure.

For the characterisation study assessing demographic and clinical profile of neonates, treatment received for the primary diagnosis and associated treatment patterns, and short-, mid-, and long-term outcomes *(objectives 2, 3(i, ii),* and *4)*, follow-up will start on the date of the first recorded diagnosis of neonatal seizure (index date), for neonates present in the data source between 1 January 2014 and 31 December 2024 (or the latest date available). End of follow-up will be defined as the earliest of: death, end of observation period (the latest available data), or study end date (31 December 2024). For treatment patterns *(objective 3(ii)),* follow-up will be restricted to the neonatal period. For assessing short-, mid-, and long-term outcomes *(objective 4),* follow-up will not be restricted to the neonatal period and will be extended to 3 months, 6 months, 2 years, and 6–7 years of age. Additionally, for outcome analyses, follow-up will end upon the first recorded outcome of interest.

For the drug utilisation study describing treatment duration among neonates receiving pharmacological therapy for seizures *(objective 3(iii)),* follow-up will start at the date of the first recorded prescription of anti-seizure medication (index date) in neonates during the study period. End of follow-up will be defined as the earliest of: death, end of observation period (the latest available data), end of treatment, or end of study period.

The code lists used to identify the start and end of follow-up determining events are presented in **ANNEX IV**.

Incidence and prevalence analyses require an appropriate denominator population and contributed observation time to first be identified. Study participants in the denominator population will begin contributing person-time-at-risk from the index date, as described in **Section 8.2.** In population-based data sources, observation begins at birth and ends at the earliest of end of the neonatal period (28 days post-birth), death, loss to follow-up/data visibility, end of data source availability, or end of study period. In hospital-based data sources, observation spans the earliest to latest recorded event, with neonatal follow-up starting at admission (if not coincident with birth) and ending at hospital discharge, end of the neonatal period, or end of observation period, whichever occurs first. Finally, in the dedicated neonatal data source, observation is from admission to discharge. For incidence analyses, follow-up additionally ends upon the first recorded diagnosis of neonatal seizure.

8.3. Study population with inclusion and exclusion criteria

For the descriptive disease epidemiology study on **incidence and prevalence** of neonatal seizures, the study population will include neonates *(objective 1):*

Inclusion criteria

- Neonates, defined as infants from birth (the date of birth (day 0)) to 28 days of life (completion of day 27).
- Present in the data source between 1 January 2014 and 31 December 2024 (or the latest available date).

Exclusion criteria

- None.

For the **characterisation study on demographic and clinical profile of neonates**, treatment received for the primary diagnosis and associated treatment patterns, and short-, mid-, and long-term outcomes among neonates diagnosed with seizures *(objectives 2, 3(i, ii),* and *4)*, the study population will include:

Inclusion criteria

- All neonates with a recorded diagnosis of seizures during the neonatal period.
- Neonatal seizure recorded in the data source between 1 January 2014 and 31 December 2024 (or the latest available date).
- No previous record of seizure diagnosis prior to the index date.
- No previous record of the outcome condition of interest prior to the index date *(objective 4).*

Exclusion criteria

- Neonates diagnosed with seizures at least <90 days prior to the end of data availability in the respective data source (to ensure sufficient follow-up) *(objective 4)*.

For the drug utilisation study describing **treatment duration** *(objective 3(iii))*, the study population will include:

Inclusion criteria

- First anti-seizure medication prescription or dispensation during the neonatal period between 1 January 2014 and 31 December 2024 (or the latest available date).
- Neonatal seizure recorded prior to the new record of anti-seizure medication prescription or dispensing.
- No prior record of anti-seizure medication prescription or dispensing record prior to the index date.

Exclusion criteria

- None.

8.4. Study setting and data sources

This study will be conducted using routinely collected data from 1 registry, 3 secondary care, and 1 neonatal unit care data sources in the DARWIN EU® network of data partners from 5 European countries, of which 4 are EU member states (**Table 1**). All data were *a priori* mapped to the OMOP CDM.

Table 1. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals*	Calendar period covered by each data source	Contributing to
Finland	FinOMOP-HUS	Inpatient (hospital) care	EHR	555,600	2014 – 08/2024	All objectives
France	CDW Bordeaux	Inpatient (hospital) care	EHR	247,200	2005 – 08/2025	All objectives
Hungary	SUCD	Inpatient (hospital) care	EHR	226,700	2005 – 03/2025	All objectives
Sweden	HI-SPEED	Registry, outpatient General Practitioner care, inpatient (hospital) care	Registry	10,563,700	2020 – 09/2024	All objectives
United Kingdom	NNRD	Neonatal units	EHR	300	2007 – 12/2021	All objectives

Abbreviations: CDW Bordeaux = Clinical Data Warehouse of Bordeaux University Hospital, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, SUCD = Semmelweis University Clinical Data, HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, NNRD = National Neonatal Research Database, EHR = Electronic Health Records. *Active individuals are defined as the maximum number of persons in an observation period, in the last 6 months of data for that source.

Data source selection

These data sources fulfil the criteria required in terms of data quality, completeness, timeliness, and representativeness for a descriptive epidemiology type of study while covering different regions of Europe (**ANNEX II**).

8.5. Study period

The study period is from 1 January 2014 to 31 December 2024, or the most recent data available for each contributing data source.

It should be noted that the availability of accurate data deviates from the start and end date of the study period in several data sources. In the HI-SPEED data source, the availability of the accurate data starts from 2020 and ends in 09/2024. Additionally, in FinOMOP-HUS and NNRD, data availability ends in 08/2024 and 12/2021, respectively.

8.6. Variables

8.6.1. Exposure

Objective 3:

In this objective, exposure to anti-seizure medication is identified as the first recorded prescription or dispensing of a specific anti-seizure drug ingredient, with no previous exposure to the same ingredient prior

to the index date. Treatment episodes will be constructed by combining sequential prescriptions into a single continuous exposure period (i.e., drug era) when the gap between the end date of one prescription and the start date of the next prescription is 3 days or less. Therefore, if two or more prescriptions are separated by no more than three days, they will be merged into one treatment episode. The exposed risk window will therefore account for all periods when the individual is likely to be using anti-seizure medication. The appropriateness of the 3-day gap and episode construction will be evaluated during cohort diagnostics. If cohort diagnostics indicate that refinements are necessary, adjustments will be documented.

The preliminary concept sets used for the identification of exposures are described in **ANNEX IV**. These codes will be refined during the study execution following the DARWIN EU® phenotyping standard processes, which involve the review of code lists and the review of phenotypes after their execution in the participating data sources. Final concept sets will be defined following input from experts from the DARWIN EU® network and EMA.

8.6.2. Outcome

The outcomes are as follows:

- Incidence and prevalence of seizures among neonates *(objective 1).*
- Characterisation of neonates at the time of seizure diagnosis *(objective 2).*
- Comorbidities and concurrent conditions occurrence in neonates at the time of seizure diagnosis *(objective 2)*.
- Prescribing of concurrent medication in neonates at the time of seizure diagnosis *(objective 2)*.
- Prescribing of treatment received for primary diagnosis of seizures in neonates *(objective 3).*
- Treatment pattern following the primary diagnosis of seizures in neonates *(objective 3).*
- Duration of anti-seizure treatment in neonates *(objective 3).*
- Death, neurodevelopmental and neurological conditions *(objective 4).* The preliminary list of outcomes of interest is listed below:
    - Death
    - Cerebral palsy
    - Developmental delay
    - Intellectual disability
    - Behavioural disorders (including attention deficit hyperactivity disorder and autism spectrum disorder)
    - Neurosensory sequelae (including hearing loss and blindness)
    - Epilepsies and, where feasible, type and/or syndromes of epilepsy (Ohtahara syndrome, West syndrome, Lennox-Gastaut syndrome, KCNQ2-related encephalopathy, early myoclonic epilepsy, epilepsy of infancy, self-limited infantile epilepsy)

The preliminary concept sets used for the identification of outcomes are described in **ANNEX IV**. These codes will be refined during the study execution following the DARWIN EU® phenotyping standard processes, which involve the review of code lists, and the review of phenotypes after their execution in the participating data sources. Final concept sets will be defined following input from experts from the DARWIN EU® network and EMA.

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

The following covariates will be used for calculation of incidence rates and prevalence of neonatal seizures among neonates *(objective 1)*:

- Calendar year
- Sex (all, males, females)
- Gestational age at birth (term (at completed 37 weeks of gestation or later), preterm (before completed 37 weeks of gestation))

If structured gestational age at birth is not available in the data sources, proxies will be applied.

The following variables will be used to characterise neonates with a record of seizure diagnosis during the neonatal period *(objective 2)*:

- Sex (males, females)
    - Assessment window at the index date: [0, 0]
- Age at diagnosis of neonatal seizures (days)
    - Assessment window at the index date: [0,0]
- Gestational age at birth and birth weight
    - Assessment window any time prior to the index date: [-Inf, 0]
- Pre-specified clinical manifestations of seizures including[3]:
    - Clonic: repetitive jerking, symmetric or asymmetric, usually of the same muscle groups
    - Myoclonic: sudden, brief involuntary muscle contractions (axial, proximal, distal)
    - Tonic: sustained muscle contraction lasting seconds to minutes
    - Autonomic changes: apnoea, irregular breathing, tachycardia, bradycardia
    - Behavioural arrest: pause of activities, freezing, unresponsiveness, blank stare
    - Automatisms: repetitive oral-facial-lingual movements (chewing, tongue thrusting, lip smacking)
    - Assessment window at the index date and 7 days after the index date: [0, 0], [0, 7]
- Diagnostic tools (e.g., EEG). Where feasible, the availability of EEG modalities, such as conventional EEG, amplitude-integrated EEG, or video EEG, will be described.
    - Assessment window any time before to 1 day before the index date, at the index date, and 30 days after the index date: [-Inf, -1], [0, 0], [1, 30]
- Concurrent pre-specified conditions including:
    - Febrile convulsions
    - Hypoxic-ischaemic encephalopathy
    - Stroke or haemorrhage
    - Infections (meningitis, sepsis, herpes simplex virus, encephalitis)
    - Traumatic brain injury

- Cortical/cerebral malformations:
  - Disorders of proliferation (hemimegalencephaly/megalencephaly, microcephaly)
  - Neuronal migration (lissencephaly, pachygyria, periventricular nodular heterotopia, schizencephaly)
  - Cortical organisation (focal cortical dysplasia, polymicrogyria, cobblestone cortex),
  - Midline/commissural defects (corpus callosum agenesis/dysgenesis, holoprosencephaly; optionally septo-optic dysplasia).

  Given potential low frequencies and variable coding granularity, a composite outcome will be analysed.
- Errors of metabolism (electrolyte imbalance, including hypoglycaemia, hypocalcaemia, hypomagnesaemia, hyponatraemia and glucose transporter protein type 1 deficiency syndrome, X-linked creatine deficiency, phenylketonuria, maple syrup urine disease, pyridoxine-dependent epilepsy, molybdenum-cofactor deficiency, hypophosphatasia, neuronal ceroid lipofuscinosis (CLN10)).
- Genetic aetiologies
  - Assessment window any time prior to 1 day before the index date, at the index date, and 30 days after the index date: [-Inf, -1], [0 , 0], [1, 30]

- Comorbidities: defined as the top 10 most frequent diagnostic codes
  - Assessment window any time prior to 1 day before the index date and at the index date: [-Inf, -1], [0 , 0]
- Comedications: described as the top 10 most prescribed medications
  - Assessment window at the index date and any time prior to the index date: [-Inf, -1], [0, 0]

Gestational age will be used for stratification of all analyses, where available.

The variables considered for characterisation of treatments received from the primary diagnosis of seizures in neonates[3] and treatment patterns analysis *(objectives 3(i, ii))* will include:

- Phenobarbital
- Phenytoin/Fosphenytoin
- Levetiracetam
- Midazolam
- Lidocaine
- Carbamazepine
- Pyridoxine
- Any additional treatments identified through the cohort diagnostics performed using the R package.
- Therapeutic hypothermia (only in *objective 3(i))*
  - Assessment window for characterisation of treatments will include index date and 30 days after index date: [0, 0], [1, 30].

Gestational age will be used for stratification of all analyses, where available.

The preliminary concept sets used for the identification of covariates are described in **ANNEX IV**. These codes will be refined during the study execution following the DARWIN EU® phenotyping standard processes, which involve the review of code lists and the review of phenotypes after their execution in the participating data sources. Furthermore, additional medications may be included based on the findings from cohort diagnostics. Specifically, we will explore which medications are administered within 30 days following a seizure diagnosis. These findings will be incorporated into the treatment characterisation and treatment pattern analyses.

8.7. Study size

No sample size has been calculated, as this is a descriptive disease epidemiology study which will not test a specific hypothesis. In addition, we will use already collected available data to estimate incidence and prevalence of seizures in neonates and characterise clinical presentation and treatment patterns in this population. Thus, the sample size is driven by the availability of data for neonates. Based on a preliminary feasibility assessment, the expected number of person counts for neonatal seizures in the data sources included in this study ranges from 600 (CDW Bordeaux) to 3,300 (HI-SPEED).

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. Operational and reporting considerations**), 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 RStudio 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

Objective 1

Incidence rates and prevalence will be calculated based on OMOP CDM mapped data using the R package *IncidencePrevalence*, developed by DARWIN EU® (https://darwin-eu.github.io/IncidencePrevalence/).

Incidence calculations

Annual and overall incidence rates will be calculated as the number of neonates with a **first recorded diagnosis** of seizure during the neonatal period, per 1,000 person-years, if appropriate (or 1,000 person-days), among the population at risk. Incidence rates will be reported for each calendar year and for the overall study period. The population at risk includes all neonates who contribute person-time from study entry until one of the following events: first recorded diagnosis of neonatal seizures (event of interest), end of neonatal period, or death (censoring at date of death). Each neonate contributes time at risk until the earliest of these events. Those who do not experience a seizure diagnosis during the neonatal period

contribute their full eligible time. Incidence rates will be accompanied by 95% confidence intervals, calculated using the Poisson distribution.

An illustration of the calculation of incidence rates for a pre-specified condition of interest is shown below in **Figure 2**. Neonate ID 1 and 4 contribute time at risk up to the point at which they are diagnosed with neonatal seizures and are considered incident cases. Neonates ID 2 and 5 do not receive a diagnosis of seizures during the neonatal period and so contribute time at risk but no incident outcomes. Meanwhile, neonate ID 3 has a recorded diagnosis of neonatal seizures prior to the start of the study period. As a result, this individual does not contribute time at risk.

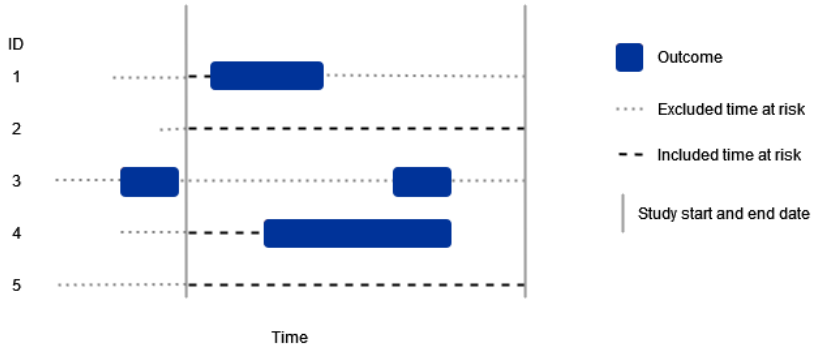


Figure 2. Incidence example.

Prevalence calculations

Prevalence of seizures among neonates will be calculated as both annual and overall period prevalence. These estimates will be calculated by dividing the total number of neonates with a recorded diagnosis of seizure during a given year or across the entire study period by the population at risk, defined as all neonates present in the data source during the corresponding time frame. Therefore, annual and overall period prevalence will reflect the proportion of neonates diagnosed with seizures at any time point during the specified interval. Binomial 95% confidence intervals will be calculated.

An illustration of the calculation of period prevalence is shown below in **Figure 3.** Between time t+2 and t+3, two of the five study participants have a record of neonatal seizure, giving a prevalence of 40%. Meanwhile, for the period t to t+1 all five also have some observation time during the year, with one of the five study participants having a recorded diagnosis of neonatal seizures, giving a prevalence of 20%.

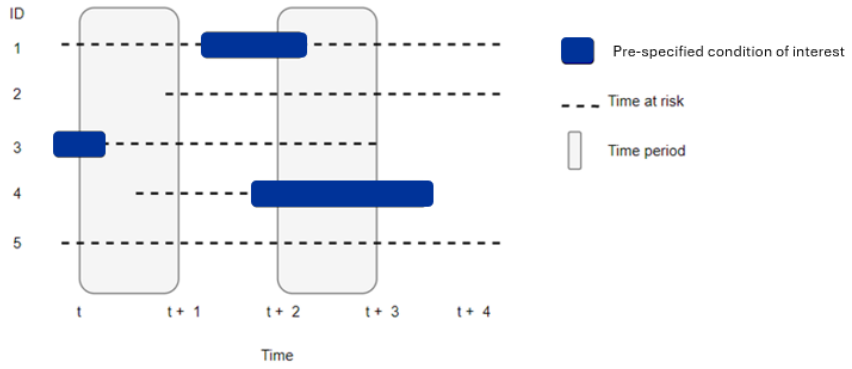


Figure 3. Period prevalence example.

Incidence rates and prevalence estimates will be presented overall, stratified by sex and gestational age (if feasible).

Objective 2

Characteristics will be described by means of large-scale characterisation and pre-specified patient-level characterisation based on OMOP CDM mapped data using the *CohortCharacteristics* R package (https://github.com/darwin-eu/CohortCharacteristics/), developed by DARWIN EU®.

Characteristics of interest will be reported as counts and proportions for categorical variables and as medians with interquartile ranges (IQR) for continuous variables. Demographic characteristics, including age at diagnosis and sex, will be assessed at the date of first recorded diagnosis of seizures in neonates (index date). Characterisation of gestational age and birth weight will be assessed any time prior to the index date. Large-scale characterisation, including characterisation of comorbidities and comedication will be assessed anytime to 1 day before the index date, and at the index date. Frequency of pre-specified symptoms and conditions, such as hypoxic-ischaemic encephalopathy, stroke or haemorrhage, infections, cortical malformations, errors of metabolism (acute and inborn), and genetic aetiologies, will be assessed anytime to 1 day before the index date, at the index date, and 30 days after the index date. Availability of diagnostic tools (EEG, video EEG) will also be assessed anytime to 1 day before the index date, at the index date, and 30 days after the index date.

Objective 3

Description of treatment received for the primary diagnosis, treatment patterns, and treatment duration will be calculated based on OMOP CDM mapped data using the *CohortCharacteristics* (https://github.com/darwin-eu/CohortCharacteristics), *DrugUtilisation* (https://github.com/darwin-eu/DrugUtilisation), and *TreatmentPatterns* (https://github.com/darwin-eu-dev/TreatmentPatterns) R packages, developed by DARWIN EU®.

Characterisation of pre-specified treatments received for the primary diagnosis of seizures

The number and percentage of neonates receiving pre-specified treatments for the primary diagnosis of seizures will be calculated. Estimates will be presented at the index date and 1 to 30 days after the index date.

Treatment pattern

The number and percentage of neonates receiving each of the pre-specified treatment options, as well as treatment combinations, will be described in the study population of neonates diagnosed with seizures (*objective 3*). Additionally, Sunburst plots and Sankey diagrams will be used to visualise treatment patterns

and sequences over time. Diagrams will be censored at the end of follow-up or death as described in **Section 8.2**.

To construct treatment pathways, various parameters can be defined in the *TreatmentPatterns* package (**Figure 4**).

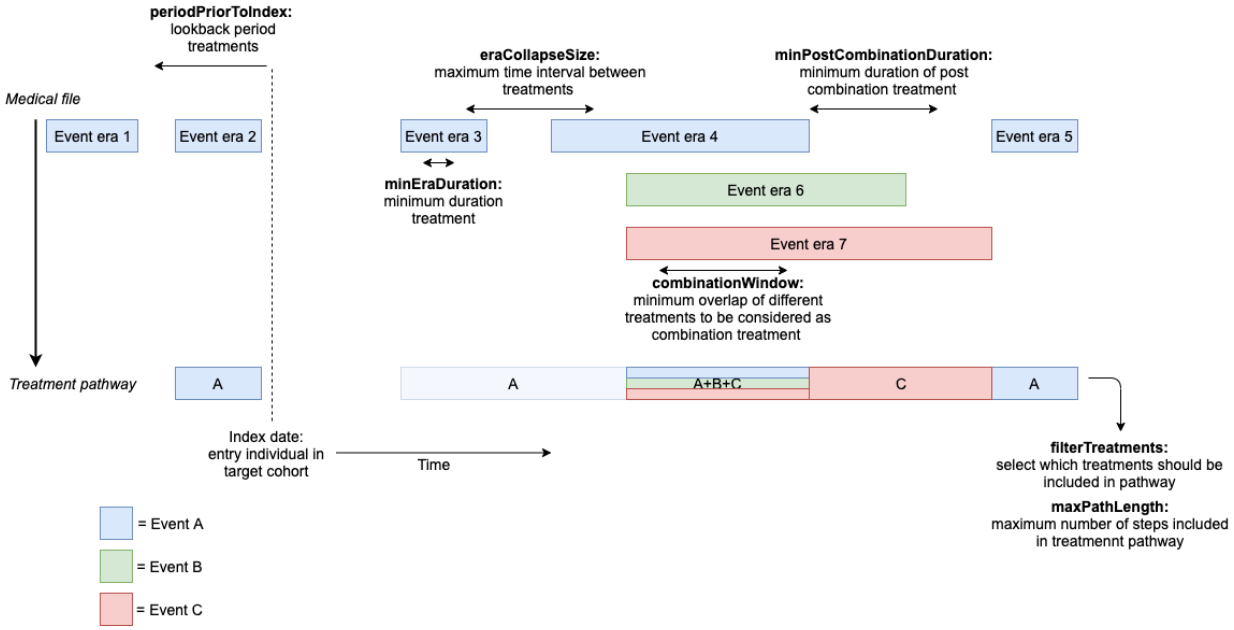


Figure 4. Parameters in *TreatmentPatterns* package.

The preliminary parameters outlined in this study are described in **Table 2**. The appropriateness of thresholds will be assessed and reevaluated during cohort diagnostics.

Table *2*. List of preliminary pathway settings with description and expected input.

Parameter	Description	Used setting
startAnchor	Reference point for the start date of the analysis.	startDate
windowStart	Number of days offset from the startAnchor to define the start of the observation window.	0
endAnchor	Reference point for the end date of the analysis.	endDate (end of neonatal period)
windowEnd	Number of days offset from the endAnchor to define the end of the observation window.	0
minEraDuration	Minimum duration (in days) that an event era must last to be included in the analysis.	1 day
splitEventCohorts	Specifies which event cohorts should be split into acute (< X days) and therapy (≥ X days) phases.	NULL
splitTime	Threshold (in days) used to divide event cohorts into acute and therapy phases.	NULL
eraCollapseSize	Time window (in days) within which two eras of the same event cohort are merged into a single era.	3 days
combinationWindow	Time window (in days) during which two overlapping event cohorts are considered a combination treatment.	3 days
minPostCombinationDuration	Minimum duration (in days) an event era must last before or after a combination treatment to be included in the analysis.	1 day

Parameter	Description	Used setting
filterTreatments	Selects changes between event cohorts	Changes
maxPathLength	Maximum number of steps allowed in a treatment pathway.	5
overlapMethod	Method for handling minor overlaps between events. "truncate" shortens the first event to align with the start of the next event.	truncate
concatTargets	Indicates whether multiple target cohorts for the same individual should be merged into a single sequence.	TRUE

Treatment duration

Treatment duration will be calculated, using the *DrugUtilisation* R package, as the duration of pre-specified anti-seizure treatment among neonates with a recorded diagnosis code for seizures during the neonatal period. Treatment duration will be summarised providing the minimum, quartiles, and maximum values. For data sources where duration cannot be calculated due to, e.g., missing information on quantity or dosing, treatment duration will not be provided.

Objective 4

Short-, mid-, and long-term outcomes among neonates diagnosed with seizures will be described based on OMOP CDM mapped data using the *CohortSurvival* (https://darwin-eu-dev.github.io/CohortSurvival/) R package, developed by DARWIN EU®.

Overall survival will be calculated using data on time at risk of death from any cause and the Kaplan-Meier (KM) method. Results will be reported as plots of the estimated survival curves, as well as the estimated probability of survival at 3 months, 6 months, 2 years, and 6–7 years of age. The KM approach implicitly assumes censoring occurs at random. This analysis will be conducted for data sources that systematically collected data on mortality. Outcomes, including related diagnosed conditions, such as neurodevelopmental and neurological disorders, will be reported as cumulative incidence plots at 3 months, 6 months, 2 years, and 6–7 years of age.

Method to deal with missing data

We assume that the absence of a prescription record or diagnostic code in the data source indicates that the individual does not receive the respective treatment or is not diagnosed with the condition of interest. Similarly, for the assessment of comorbidities, we assume that the absence of a recorded diagnostic code for a given condition means that that condition is not present or not recorded in the context of routine clinical care.

Sensitivity analysis

None.

8.8.4. Output

Output will include the following:

A PDF report including an executive summary and the following tables and figures.

Table 1. Mock table shell – Study attrition of neonates based on diagnosis of seizures during the neonatal period, presented by data source.

	FinOMOP-HUS	CDW Bordeaux	SUCD	HI-SPEED	NNRD
Qualifying initial records[1]					
Cohort start date between 1/1/2014 and 31/12/2024					
Neonatal seizure record					
Potential 90 days of follow-up					
No previous record of seizures					

[1]Refers to infants from birth (the date of birth (day 0) to 28 days of life (completion of day 27), the study period for individuals' inclusion (2014–2024).

Table 2. Mock table shell – Overall incidence rate of seizures in neonates, by data source, during 2014–2024 *(objective 1)*.

Data source	Number of persons	Number of events	Person-years (PYs)	Incidence per 1,000 PYs	95% CI
FinOMOP-HUS					
CDW Bordeaux					
SUCD					
HI-SPEED					
NNRD					

Table 3. Mock table shell – Overall incidence rate of seizures in neonates, stratified by sex and data source *(objective 1)*.

Data source	Sex	Number of persons	Number of events	Person-years (PYs)	Incidence per 1,000 PYs	95% CI
FinOMOP-HUS	Male					
	Female					
CDW Bordeaux	Male					
	Female					
SUCD	Male					
	Female					
HI-SPEED	Male					
	Female					
NNRD	Male					
	Female					

Table 4. Mock table shell – Overall prevalence of seizures in neonates, by data source *(objective 1)*.

Data source	Number of persons	Number of events	Prevalence	95% CI
FinOMOP-HUS				
CDW Bordeaux				
SUCD				
HI-SPEED				
NNRD				

Table 5. Mock table shell – Overall prevalence of seizures in neonates, by sex and data source *(objective 1)*.

Data source	Sex	Number of persons	Number of events	Prevalence	95% CI
FinOMOP-HUS	Male				
	Female				
CDW Bordeaux	Male				
	Female				
SUCD	Male				
	Female				
HI-SPEED	Male				
	Female				
NNRD	Male				
	Female				

Similar output tables will be generated stratified by gestational age, where available.

Table 6. Mock table shell – Demographic and clinical characteristics of neonates with recorded seizure diagnosis during the neonatal period *(objective 2)*.

Variable	Format	FinOMOP-HUS	CDW Bordeaux	SUCD	HI-SPEED	NNRD
Number of subjects	N					
Number of records	N					
Cohort start date	Median [q25 – q75]					
Cohort end date	Median [q25 – q75]					
Prior observation	Median [q25 – q75]					
Future observation	Median [q25 – q75]					

Variable	Format	FinOMOP-HUS	CDW Bordeaux	SUCD	HI-SPEED	NNRD
Age	Median [q25 – q75]					
Sex						
• Male	N (%)					
• Female	N (%)					
Gestational age at birth	Median [q25 – q75]					
Clinical manifestations						
• Clonic	N (%)					
• Myoclonic	N (%)					
• Tonic	N (%)					
• Autonomic changes	N (%)					
• Behavioural arrest	N (%)					
• Automatisms	N (%)					
Diagnostic tools						
EEG	N (%)					
• Conventional EEG	N (%)					
• Amplitude EEG	N (%)					
• Video EEG	N (%)					
Concurrent pre-specified conditions including						
• Febrile convulsions	N (%)					
• Hypoxic-ischaemic encephalopathy	N (%)					

Variable	Format	FinOMOP-HUS	CDW Bordeaux	SUCD	HI-SPEED	NNRD
• Stroke or haemorrhage	N (%)					
• Infections (meningitis, sepsis, herpes simplex virus)	N (%)					
• Traumatic brain injury	N (%)					
• Cortical/cerebral malformations	N (%)					
• Errors of metabolism	N (%)					
• Genetic aetiologies	N (%)					

Table 7. Mock table shell – Top 10 recorded conditions in neonates with recorded seizure diagnosis, categorised by data source, at the index date *(objective 2)*.

FinOMOP-HUS		CDW Bordeaux		SUCD		HI-SPEED		NNRD	
Condition	n (%)	Condition	n (%)	Condition	n (%)	Condition	n (%)	Condition	n (%)

Table 8. Mock table shell – Top 10 recorded conditions in neonates with recorded seizure diagnosis, categorised by data source, any time prior to the index date *(objective 2)*.

FinOMOP-HUS		CDW Bordeaux		SUCD		HI-SPEED		NNRD	
Condition	n (%)	Condition	n (%)	Condition	n (%)	Condition	n (%)	Condition	n (%)

FinOMOP-HUS		CDW Bordeaux		SUCD		HI-SPEED		NNRD	
Condition	n (%)	Condition	n (%)	Condition	n (%)	Condition	n (%)	Condition	n (%)

Table 9. Mock table shell – Top 10 most prescribed medications in neonates with recorded seizure diagnosis, categorised by data source, at the index date *(objective 2)*.

FinOMOP-HUS		CDW Bordeaux		SUCD		HI-SPEED		NNRD	
Medication	n (%)	Medication	n (%)	Medication	n (%)	Medication	n (%)	Medication	n (%)

Table 10. Mock table shell – Top 10 most prescribed medications in neonates with recorded seizure diagnosis, categorised by data source, any time prior to the index date *(objective 2)*.

FinOMOP-HUS		CDW Bordeaux		SUCD		HI-SPEED		NNRD	
Medication	n (%)	Medication	n (%)	Medication	n (%)	Medication	n (%)	Medication	n (%)

Table 11. Treatments received for the primary diagnosis of seizures in neonates at the index date.

Treatment		FinOMOP-HUS	CDW Bordeaux	SUCD	HI-SPEED	NNRD
Phenobarbital	N (%)					
Phenytoin	N (%)					
Levetiracetam	N (%)					
Midazolam	N (%)					
Lidocaine	N (%)					
Carbamazepine	N (%)					
Medication x	N (%)					
Medication y	N (%)					

Table 12. Treatment duration of pre-specified medication of interest for neonatal seizures during the study period and across the data sources.

Treatment		FinOMOP-HUS	CDW Bordeaux	SUCD	HI-SPEED	NNRD
Phenobarbital	Median, (q25, q75)					
	Range (min – max)					
Phenytoin	Median, (q25, q75)					
	Range (min – max)					
Levetiracetam	Median, (q25, q75)					
	Range (min – max)					
Midazolam	Median, (q25, q75)					
	Range (min – max)					
Lidocaine	Median, (q25, q75)					
	Range (min – max)					
Carbamazepine	Median, (q25, q75)					
	Range (min – max)					
Medication x	Median, (q25, q75)					
	Range (min – max)					
Medication y	Median, (q25, q75)					
	Range (min – max)					

Similar output tables will be generated stratified by gestational age, where available.

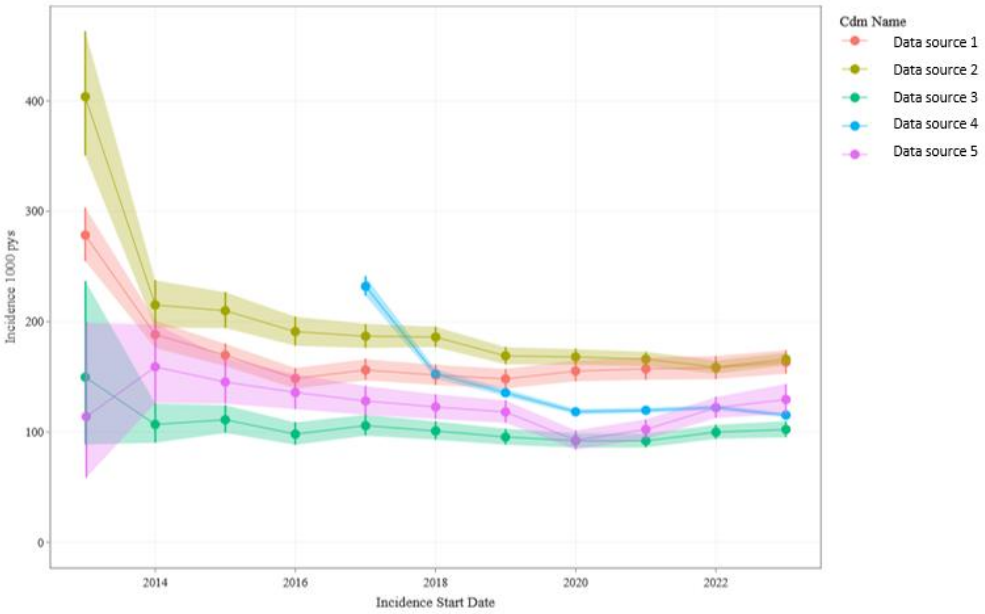


Figure 1. Mock figure shell – Annual incidence rate of seizures in neonates, by data source *(objective 1)*.

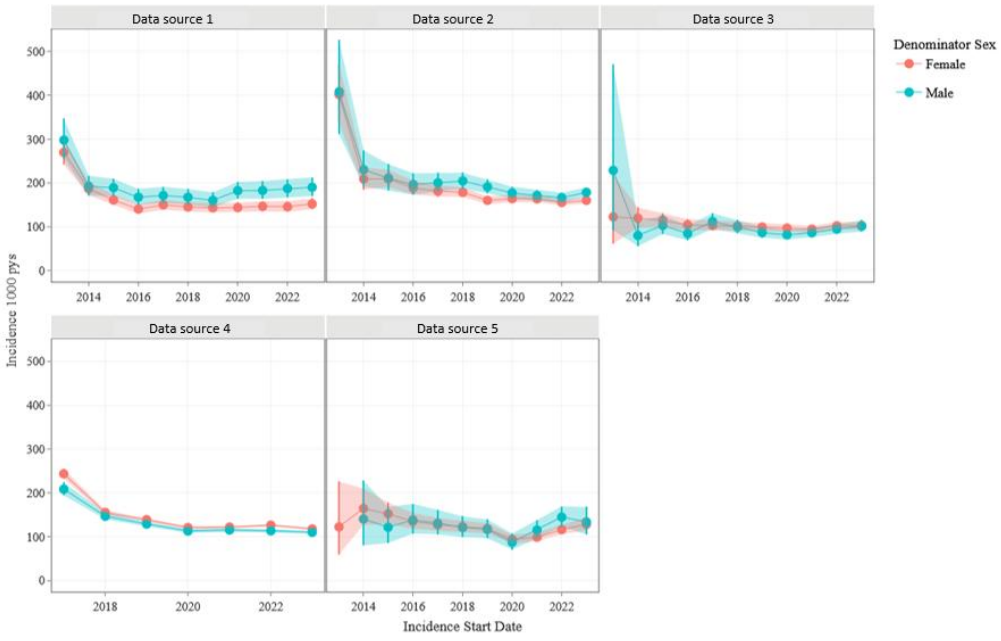


Figure 2. Mock figure shell – Annual incidence rate of seizures in neonates, stratified by sex and data source *(objective 1)*.

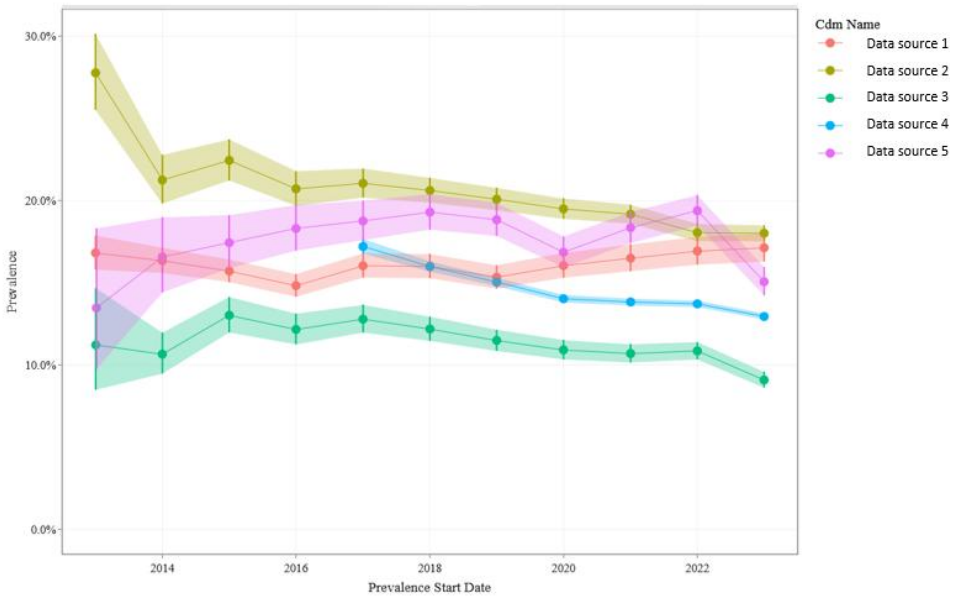


Figure 3. Mock figure shell – Annual prevalence of seizures in neonates, by data source *(objective 1)*.

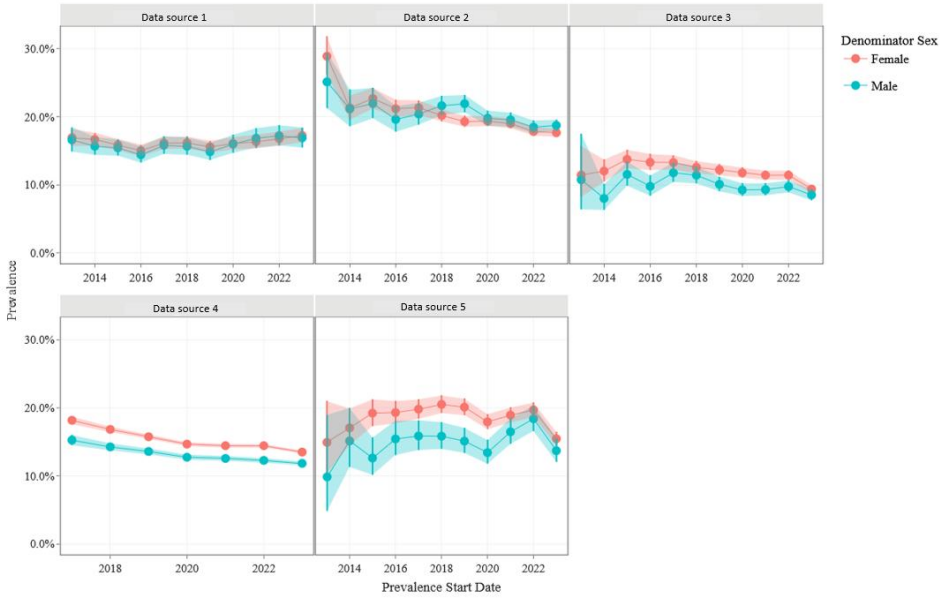


Figure 4. Mock figure shell – Annual prevalence of seizures in neonates, stratified by sex and data source *(objective 1)*.

Similar output figures will be generated stratified by gestational age, where available.

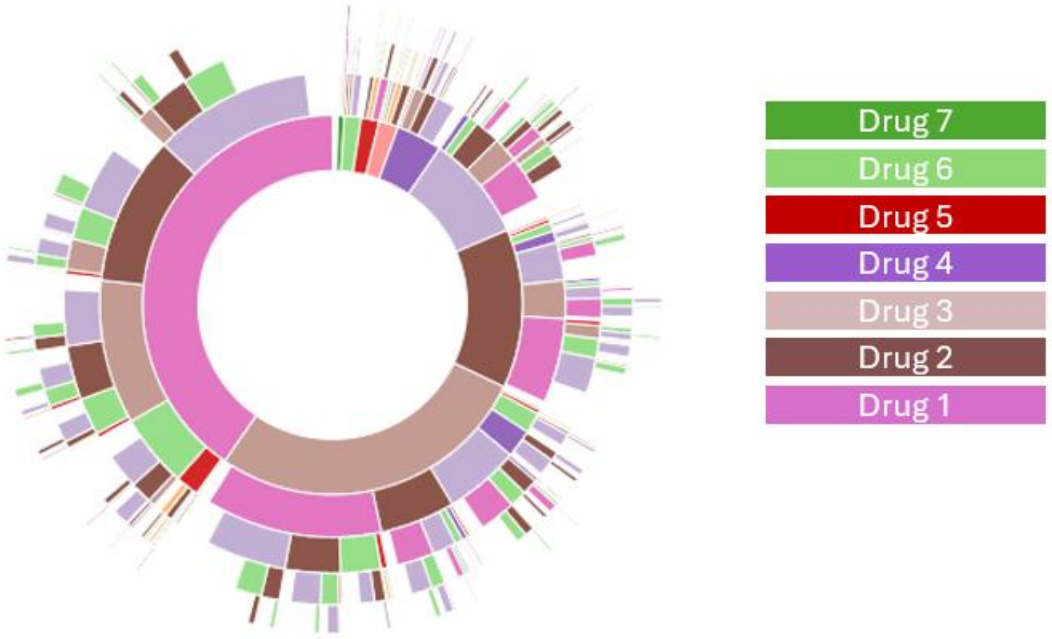


Figure 5. Sunburst plot depicting treatment patterns following seizure diagnosis in neonates, overall in data source 1 *(objective 3).*

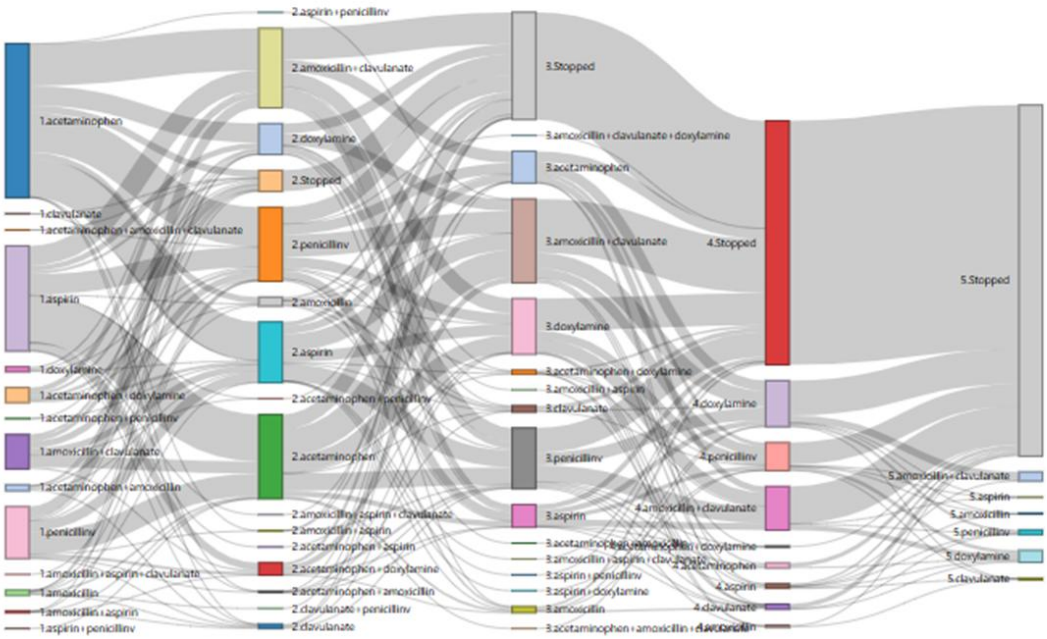


Figure 6. Sankey plot depicting treatment sequence following seizure diagnosis in neonates, overall in data source 1 *(objective 3).*

Figures 5 and 6 are provided as mock visualisations to illustrate the intended graphical format. Similar Sunburst and Sankey plots will be generated for each participating data source.

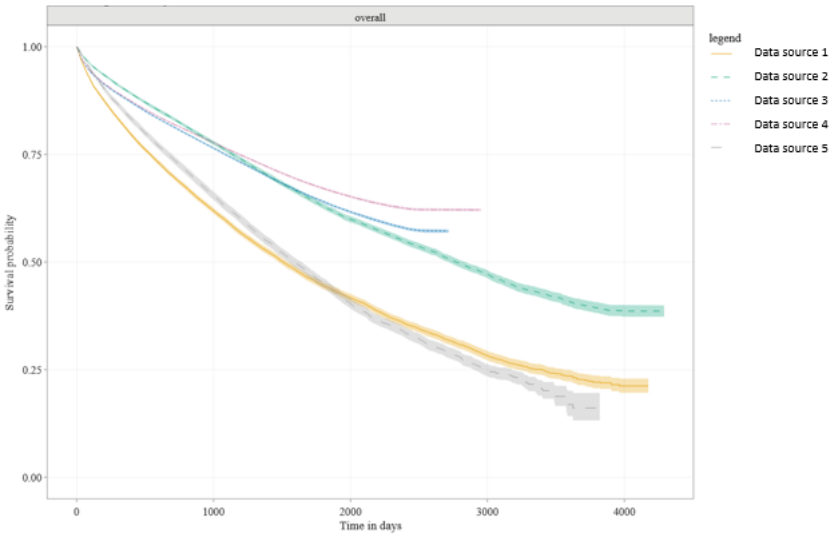


Figure 7. Survival probability of neonates with seizures, at 3 months, 6 months, 2 years, and 6–7 years of age, by data source.

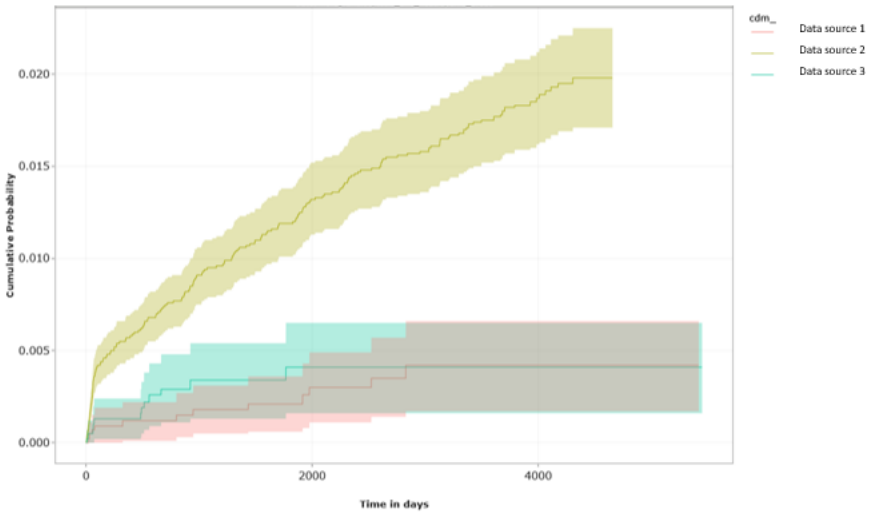


Figure 8. Cumulative incidence of epilepsy at 3 months, 6 months, 2 years, and 6–7 years of age.

Figure 9. Cumulative incidence of cerebral palsy at 3 months, 6 months, 2 years, and 6–7 years of age.

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

8.9. Evidence synthesis

Results from analyses described in **Section 8.8.3.** will be presented separately for each data source. No meta-analysis of results will be conducted.

9. STRENGTHS AND LIMITATIONS

Strengths

This study leverages a large and diverse set of real-world data sources from five European countries, including Finland, France, Hungary, Sweden, and the United Kingdom, covering registry-based, inpatient hospital care, and neonatal care data sources. The broad geographic coverage and nature of these data sources enable an updated assessment of neonatal seizures across varied healthcare settings. Use of the OMOP CDM ensures standardised data structuring, harmonised variables, and consistent cohort definitions across sources. This facilitates reproducible analyses and enables robust characterisation of demographic and clinical profiles, treatment patterns, and outcomes in neonates diagnosed with seizures.

Limitations

This study will be informed by routinely collected healthcare data, which introduces several important considerations that may influence the interpretation of the findings.

The study population is limited to neonates captured within the participating data sources, and findings may not be generalisable to other countries or healthcare systems with different clinical practices or data capture mechanisms. For hospital-based data sources, incidence and prevalence estimates will reflect the risk among hospitalised neonates only and may not represent the broader neonatal population.

The diagnosis of neonatal seizures is challenging and typically requires diagnostic tools, such as EEG. As these diagnostic results are not available in the data sources used for this study, identification of neonatal seizures will rely solely on recorded diagnosis codes, which may reduce accuracy and introduce misclassification bias. Furthermore, no prior validation of neonatal seizure identification using diagnosis codes alone has been conducted in these data sources.

Electronic health records and registries are primarily designed for clinical and administrative purposes rather than research. As such, documentation of key variables, including symptoms, diagnostic procedures (e.g., conventional EEG, amplitude EEG, or video EEG), and comorbidities, may be incomplete, inconsistent, or variably recorded across sources. This may affect the accuracy of patient-level characterisation.

Gestational age at birth and birth weight, which are essential for neonatal profiling, may be missing or inconsistently recorded or not mapped in some data sources. Similarly, identification of high-risk conditions (e.g., hypoxic-ischaemic encephalopathy, stroke, infections, genetic aetiologies) depends on the availability and granularity of diagnostic coding.

A recorded drug prescription does not necessarily indicate that the individual actually took the drug. Therefore, assumptions of actual use were made.

Treatment duration and patterns are subject to limitations due to gaps in medication documentation. End-of-treatment dates are not consistently available, and imputation methods using fixed duration assumptions (aligned with OMOP CDM conventions) may be applied during the ETL process. While this promotes consistency, it may not fully reflect true treatment variability.

Outcome assessment at 3 months, 6 months, 2 years, and 6–7 years of age depends on longitudinal data availability and completeness. Loss to follow-up or data truncation may affect the reliability of long-term outcome estimates, including survival and neurodevelopmental conditions. In SUCD, mortality data from this hospital source is restricted to in-hospital deaths only, there is no linkage to national death registries.

Consequently, overall mortality estimates cannot be derived. Analyses will be limited to in-hospital mortality.

10. REFERENCES

1. Shellhaas, R.A., *Chapter 17 - Seizure classification, etiology, and management*, in *Handbook of Clinical Neurology*, L.S. de Vries and H.C. Glass, Editors. 2019, Elsevier. p. 347-361.
2. *Neonatal seizures*. [cited 2025 03/10/2025]; Available from: https://www.ucsfbenioffchildrens.org/conditions/neonatal-seizures.
3. Pressler, R.M., et al., *The ILAE classification of seizures and the epilepsies: Modification for seizures in the neonate. Position paper by the ILAE Task Force on Neonatal Seizures.* Epilepsia, 2021. **62**(3): p. 615-628.
4. Kim, E.H., J. Shin, and B.K. Lee, *Neonatal seizures: diagnostic updates based on new definition and classification.* Clin Exp Pediatr, 2022. **65**(8): p. 387-397.
5. Abiramalatha, T., et al., *Anti-seizure medications for neonates with seizures.* Cochrane Database Syst Rev, 2023. **10**(10): p. Cd014967.

11. ANNEXES

ANNEX I. Description of data sources

Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)

#	Section	Description
1	Database identification and country	FinOMOP-HUS (Hospital District of Helsinki and Uusimaa) Finland, Uusimaa
2	Data partner information section	HUS – Helsinki University Hospital, Hospital District of Helsinki and Uusimaa, HUS Tietohallinto (Information Management)
3	Coverage and timespan	Data collection since: 2014 Extent: Regional. HUS is responsible for specialised healthcare in Finland's Uusimaa region and treatment of many rare and severe diseases that are nationally centralised to HUS.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care) and hospital inpatient care. All visits, examinations, laboratory tests, procedures, and treatments are recorded in the HUS IT systems and integrated into the data lake. The data lake stores decades of clinical information in digital format, and data from both past and current source systems are available. Systems providing data into the data lake include: CGI Uranus and Epic Apotti (EHR: visits, diagnoses, medication, etc.), Opera (procedure records), Kemokur and Beacon (cancer-specific medications), Marela (hospital pharmacy), Multilab/Mylab+ (laboratory system), Qpati/Mylab+ (pathology records system).
5	Data collection process	Outpatient electronic health records, inpatient hospital electronic health records, and other. All visits, all procedures, and all given treatments have been recorded systematically in the electronic format. We use more than a hundred different operational IT systems.
6	General representativeness	The data reflects patients that have visited the hospital and is thus not a generic population.
7	Data content / source coding	Drugs: ATC, generic name, brand name + dosage and strength information in the data + local route mapping Procedures: Nomesco NCSP Genetic: special mappings to Omop Genomics Diagnosis: ICD10 + Finnish adaptation “ICD10fi” Laboratory tests: National laboratory test numbers + local hus variants, National microbe listings, local “additional information” - coding Various observations during care: LOINC and local codings
8	Data Harmonisation	Complete. All patients are identified with a unique, national social security number, which is a permanent person identifier. Based on this, the derived patient identifiers (pseudonyms) in the HUS data lake are also unique.
9	Quality control (database specific)	The HUS data lake and IT management is ISO 13485:2016 and ISO 9001:2015 certified and has capabilities in developing, validating, and CE-marking medical devices and software. Data picks for individual research projects are performed by a dedicated data analyst team and are subject to internal review before data are released for OMOP mapping and into research-specific analysis environments.
10	Linkage	Data from many nationwide registries can be combined at HUS, which is subject to obtaining special data permits.

#	Section	Description
11	Vital status	
12	Limitations	
13	Main references	No main reference provided
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111134 Website: https://www.hus.fi/en

France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)

#	Section	Description
1	Database identification and country	CDW Bordeaux (Clinical Data Warehouse of Bordeaux University Hospital) France, Nouvelle-Aquitaine
2	Data partner information section	CHU DE BORDEAUX - DIRECTION GENERALE Gironde / Nouvelle-Aquitaine
3	Coverage and timespan	Data collection since: 2005 Extent: Regional. It covers the population of Bordeaux metropolitan area, and possibly beyond, as the healthcare centre for referrals and expertise for the Nouvelle Aquitaine region. The data source contains data from 2005 onwards.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), hospital inpatient care, and claims data. The data source currently captures information about patient demographics, visit details, conditions, procedures, drugs, measurements, and mortality.
5	Data collection process	Outpatient electronic health records, inpatient hospital electronic health records, inpatient hospital billing systems, and biobank. The integrated data is extracted from the hospital production information system via a real-time research database. The data is then processed and quality controlled by a team dedicated to maintaining the database. Internal evaluations were carried out to ensure consistency between the research database and the patient bedside software.
6	General representativeness	This is the 6th largest metropolitan area in France, and CHUBX is the largest hospital in the region. More than 75% of the patients admitted to Bordeaux University Hospital reside in the Gironde departments, with almost 50% coming directly from the Bordeaux metropolitan area. The hospital also captures additional cases from Nouvelle-Aquitaine region.
7	Data content / source coding	Diagnosis source data is coded in ICD-10 terminology. Procedures are coded in CCAM (French terminology). Laboratory measurements are coded in local terminology and partially mapped to LOINC. Drugs are coded through a local terminology and then mapped to UCD (French terminology), as well as ATC codes.
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 have been verified upon onboarding into the DARWIN EU® data network. We use the hospital's unique identifier to generate the patient identifier in OMOP. If two identities are merged at the hospital, the merge is taken into account in the CDW. An automatic (hourly) detection of suspected duplicated identities has been implemented at the hospital since 2020, with merging of duplicated identities by a specialised team. Identities since 2015 were

#	Section	Description
		processed retrospectively. Thus, the rate of identity duplication in the database is low, especially since 2015.
9	Quality control (database specific)	The integrated data comes from the hospital production information system through a real-time replicated database. Consistency evaluations between the replicated database and the production system are performed by the technical team in charge of maintaining the replicated database. In the same way, consistency checks are performed between the replicated database and the data integrated into the i2b2 CDW. In addition, dashboards enable monitoring the data integrated into the i2B2 CDW, in particular by controlling the amount of data available over time and its evolution, according to the various data sources. An internal evaluation was carried out to ensure the consistency between the data integrated into i2b2 and the data available in the software used at the patient's bedside. In addition, many use cases were performed on the i2b2 CDW, with return to the patient chart and comparison of the data integrated into i2b2 and the data available in the care file.
10	Linkage	Death certificates (without the cause of death).
11	Vital status	The data source is linked to the French death registry.
12	Limitations	CDW Bordeaux is limited to events captured in the hospital setting and thus does not include patient events not treated by the hospital (e.g., rare cancers). Patient events that are not included in CDW Bordeaux are rare disease treatments or specialist events that occur outside of CHUBX.
13	Main references	Cossin S, Diouf S, Griffier R, Le Barrois d'Orgeval P, Diallo G, Jouhet V "Linkage of Hospital Records and Death Certificates by a Search Engine and Machine Learning." JAMIA open (2021): 33709061
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111112 Website: https://www.chu-bordeaux.fr/

Hungary: Semmelweis University Clinical Data (SUCD)

#	Section	Description
1	Database identification and country	SUCD (Semmelweis University Clinical Data) Hungary, Budapest
2	Data partner information section	Semmelweis University
3	Coverage and timespan	Data collection since: 2005 Extent: Regional. The general catchment area of SU is the central region of the country, Budapest city and Pest County, although patients can be referred from anywhere in Hungary. The total population of Budapest and Pest County is approximately 4,200,000 people. The total population of Hungary is around 9,500,000.
4	Healthcare setting / type of data	Secondary care - specialists (ambulatory or hospital outpatient care), hospital inpatient care, claims data, and other. Diagnostic data (laboratory tests, radiology, pathology)
5	Data collection process	Insurance/administrative claims, outpatient electronic health records, inpatient hospital electronic health records, inpatient hospital billing systems, and registries. Data is extracted directly from the source database. From there, the data entry in the system is heavily controlled and validated on the user interface before being made available for further research.

#	Section	Description
6	General representativeness	SU captures information on patients who are covered by the public health insurance system. This covers all Hungarian citizens, and therefore the database should mirror the source population well.
7	Data content / source coding	Regarding SU's source data, procedures and diagnoses are coded in SNOMED, measurements are coded in LOINC, and drugs are stored in RxNorm and ATC.
8	Data Harmonisation	Complete. No, we use a SSN that is unique.
9	Quality control (database specific)	The clinical database is the source database and therefore it has to be treated as a trusted database. Data entry in the systems is heavily controlled by validation on the user interface, and there are a large number of rules that control the data on the insurer's side that have to be corrected in the system by the users to be able to close the encounters. OMOP mapping is done in the framework by EHDEN recognised partners under quality check by the EHDEN society.
10	Linkage	-
11	Vital status	
12	Limitations	
13	Main references	No main reference provided
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000184 Website: https://www.semmelweis.hu

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

#	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, 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 national and regional registries, only once all legislative, GDPR and ethical approvals have been granted. Therefore, only relevant data is passed on, which will then be entered and processed by the
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 clinical procedures.
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

#	Section	Description
		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 of data type are applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: Website:

United Kingdom: National Neonatal Research Database (NNRD)

#	Section	Description
1	Database identification and country	NNRD (National Neonatal Research Database) United Kingdom
2	Data partner information section	Imperial College London Section of Neonatal Medicine
3	Coverage and timespan	Data collection since: 2007 Extent: Nation-wide. The geographic catchment area spans the entire of England, Wales, and Isle of Man. The combined estimated population of these regions is approximately 80 million people.
4	Healthcare setting / type of data	Other (specify). Neonatal units.
5	Data collection process	Other. Patient data is received from EPRs that are extracted from NHS neonatal units. Data quality checks are carried out for minimum requirements, cleaning, and processing. The participating unit that the data was obtained from will have issues reported to them.
6	General representativeness	The neonatal data comes from all socioeconomic backgrounds and should therefore mirror the neonatal population well.
7	Data content / source coding	No code terminologies are used in the source system. Drugs, conditions, and procedures are captured in structured English text.
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. The same patients cannot be registered under a different identifier in the OMOP CDM data.
9	Quality control (database specific)	Data are received as a standard quarterly extract (monthly from 2025) from the Electronic Patient Records of all admissions to NHS neonatal units unless parent opt-out applies. An initial minimum data quality check is carried out on all records in the Episodes table to identify records with no birth year, gestational age, admission time, or negative episode number; these records are removed and held in a separate table. Episode duplicates are then removed in one of two ways: o Any records with a duplicated baby identifier and episode number are taken out of the NNRD and stored in a separate table o Any records with a duplicated baby identifier, admission time, and hospital code are taken out of the NNRD and stored in a separate table

#	Section	Description
		Further detailed quality checks are carried out on data and reported back to participating units. These checks include highlighting missing or potentially erroneous data and admission and discharge times that do not follow sequentially. Once quality checks have been applied, updates are implemented through a data synchronisation SQL script. This inserts new records into the data tables and updates new values for coexisting records, while enforcing referential integrity.
10	Linkage	There are no linked datasets.
11	Vital status	Entry and exit dates are defined by the admission and discharge dates of the patient, information on deaths is only captured when they occur within the visit to the neonatal unit.
12	Limitations	No database-specific limitations documented. General limitations of data type are applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: Website: https://www.imperial.ac.uk/neonatal-data-analysis-unit

ANNEX II. Fitness for use assessment

Data source justification for inclusion and key characteristics

Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)

FinOMOP-HUS will be included in this study because it is an inpatient (hospital) data source that provides relevant information on medication use and seizure-related outcomes in the population of neonates.

Based on a preliminary feasibility assessment, the expected number of person counts for neonatal seizures (all objectives) will be 1,200.

Moreover, data availability and follow-up in FinOMOP-HUS is sufficient, as data availability starts in 2014, and the date of the most recent data extraction is 08/2024, which aligns with the study period. The median follow-up of the first observation period is 2,470 days.

There are some study-specific limitations present in FinOMOP-HUS, namely, due to data availability, the study period for FinOMOP-HUS will be from 2014 to 08/2024.

Lastly, IRB approval for FinOMOP-HUS is estimated to take 1 to 3 months, which makes the execution of this study feasible within the current study timelines.

France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)

CDW Bordeaux will be included in this study because it is an inpatient (hospital) data source that provides relevant information on medication use and seizure-related outcomes in the population of neonates.

Based on a preliminary feasibility assessment, the expected number of person counts for neonatal seizures (all objectives) will be 600.

Moreover, data availability and follow-up in CDW Bordeaux are sufficient, as data availability in 2005, and the date of most recent data extraction is 08/2025, which aligns with the study period. The median follow-up of the first observation period is 247 days.

There are no study-specific limitations present in CDW Bordeaux.

Lastly, IRB approval is estimated to take 1 to 2 weeks, which makes the execution of this study feasible within the current study timelines.

Hungary: Semmelweis University Clinical Data (SUCD)

SUCD will be included in this study because it is an inpatient (hospital) data source that provides relevant information on medication use and seizure-related outcomes in the population of neonates.

Based on a preliminary feasibility assessment, the expected number of person counts for neonatal seizures (all objectives) will be 1,500.

Moreover, data availability and follow-up in SUCD are sufficient, as data availability starts in 2005, and the date of the most recent data extraction is 03/2025, which aligns with the study period. The median follow-up of the first observation period is 243 days.

There are no study-specific limitations present in SUCD.

Lastly, IRB approval is estimated to take 1 to 3 months, which makes the execution of this study feasible within the current study timelines.

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

HI-SPEED will be included in this study because it is a registry, outpatient general practitioner, and inpatient (hospital) data source that provides relevant information on medication use and seizure-related outcomes in the population of neonates.

Based on a preliminary feasibility assessment, the expected number of person counts for neonatal seizures (all objectives) will be 3,300.

Moreover, data availability and follow-up in HI-SPEED are sufficient, as data availability starts in 2020, and the date of most recent data extraction is 09/2024. The median follow-up of the first observation period is 3,530 days.

There are some study specific limitations present in HI-SPEED, namely, due to data availability, the study period for HI-SPEED will be from 2020 to 09/2024.

Lastly, IRB approval is estimated to take up to 1 month, which makes the execution of this study feasible within the current study timelines.

United Kingdom: National Neonatal Research Database (NNRD)

NNRD will be included in this study because it is a neonatal data source that provides relevant information on medication use and seizure-related outcomes in the population of neonates.

Based on a preliminary feasibility assessment, the expected number of person counts for neonatal seizures (all objectives) will be 800.

Moreover, data availability and follow-up in NNRD are sufficient, as data availability starts in 2007, and the date of the most recent data extraction is 12/2021, which aligns with the study period. The median follow-up of the first observation period in NNRD is 4 days.

There are some study-specific limitations present in NNRD, namely, due to data availability, the study period for NNRD will be from 2014 to 12/2021.

Lastly, IRB approval is estimated to take up to 3 months, which makes the execution of this study feasible within the current study timelines.

Table 1. Fitness-for-use assessment of data sources.

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant	Suggested extensiveness assessment	Suggested assessment of other quality dimensions	Suggested substantiation by documentation
Study population	Neonates, from birth to 28 days of life (obj. 1)	Date of birth (date)	High	80% will have a date of birth available	N/A	Documentation on the RWD source's target population age range and format (if available)
	Confirmed diagnosis of neonatal seizures (obj. 2–4)	Physician diagnosis of neonatal seizures (SNOMED code)	High	100% will have a diagnosis of neonatal seizures, the expected number of persons counts for neonatal seizures in the data sources included in this study ranges from 600 (CDW Bordeaux) to 3,300 (HI-SPEED)	100% of diagnosis have been mapped to SNOMED	Documentation on mapping of different coding systems to SNOMED
Treatment/exposure	Anti-seizure treatment prescription (obj. 3)	Prescription of pre-specified anti-seizure treatment (RxNorm code)	Medium	N/A (to be assessed on a research question basis)	N/A	N/A
Comparator group (not relevant)	N/A	N/A	N/A	N/A	N/A	N/A
Key outcomes per objective	Neonatal seizures (obj. 1)	Physician diagnosis of neonatal seizures (SNOMED code or equivalent)	Medium	Neonatal seizure diagnosis is included in all data sources, the expected number of persons counts for neonatal seizures in the data sources included in this study ranges from 600 (CDW Bordeaux) to 3,300 (HI-SPEED)	N/A	N/A

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant	Suggested extensiveness assessment	Suggested assessment of other quality dimensions	Suggested substantiation by documentation
	Demographic and clinical characteristics (obj. 2)	Concept IDs for sex, age, and various clinical characteristics	Medium	N/A (to be assessed on a research question basis)	N/A	N/A
	Received anti-seizure treatment prescription (obj. 3)	Prescription of pre-specified anti-seizure treatment (RxNorm code)	Medium	N/A (to be assessed on a research question basis)	N/A	N/A
		Treatment duration (start and end date of treatment)	Medium	Completeness to be assessed during the diagnostics stage	N/A	N/A
	Overall survival (obj. 4)	Date of death (date)	Medium	Completeness to be assessed during the diagnostics stage	N/A	N/A
		Physician diagnosis of epilepsy (SNOMED code or equivalent)	Medium	Completeness to be assessed during the diagnostics stage	N/A	N/A
		Physician diagnosis of cerebral palsy and other pre-specified outcomes (SNOMED code or equivalent)	Medium	Completeness to be assessed during diagnostics stage	N/A	N/A
Covariates (including confounders if relevant)	N/A	N/A	N/A	N/A	N/A	N/A
Follow-up time (if relevant)	N/A	N/A	N/A	N/A	N/A	N/A

EMA Data Quality Framework for EU medicines regulation: application to Real-World Data for more information (https://www.ema.europa.eu/system/files/documents/other/data-quality-framework-eu-medicines-regulation-application-real-world-data_en.pdf).

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 are 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 standardised 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 drug cohorts, non-systemic products will be excluded from the list of included codes summarised on the ingredient level.

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, *DrugUtilisation* to characterise the drug use, *CohortCharacteristics* to characterise the cohort, *TreatmentPatterns* to characterise treatment patterns, and *CohortSurvival* to assess short-, medium- and long-term outcomes. 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® 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: List of stand-alone documents

Table S1. Preliminary list of conditions definitions.

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Neonatal seizures	Neonatal seizure	380533	-	SNOMED
Neonatal seizures	Epileptic seizure	4197485	-	SNOMED
Neonatal seizures	Simple partial epileptic seizure	44789005	-	SNOMED
Neonatal seizures	Epileptic seizure witnessed by history provider	37153689	-	SNOMED
Gestational age	Gestational age--at birth	46234792	-	LOINC
Gestational age	Estimated fetal gestational age at delivery	40485048	-	SNOMED
Gestational age	Finding of gestational age	44792185	-	SNOMED
Gestational age	Gestational age	1077378	-	SNOMED
Birth weight	Weight of neonate at birth	3662222	-	SNOMED
Birth weight	Birth weight	4264825	-	SNOMED
Birth weight	Birth weight finding	4187520	-	SNOMED
Birth weight	Birth weight Measured	3011043	-	SNOMED
Clinical manifestations				
Clonic seizures	Neonatal focal clonic epileptic seizure	1244425	-	SNOMED
Clonic seizures	Clonic epileptic seizure	4104853	-	SNOMED
Myoclonic seizures	Neonatal focal myoclonic epileptic seizure	1244423	-	SNOMED
Myoclonic seizures	Myoclonic epileptic seizure	37163355	-	SNOMED
Tonic seizures	Tonic epileptic seizure	4102344	-	SNOMED
Tonic seizures	Neonatal focal tonic epileptic seizure	1244424	-	SNOMED
Autonomic seizures	Neonatal focal autonomic epileptic seizure	1244427	-	SNOMED
Autonomic seizures	Focal onset autonomic epileptic seizure	37168480	-	SNOMED
Behavioural arrest	Neonatal focal behavioral arrest epileptic seizure	1244428	-	SNOMED
Behavioural arrest	Behavioral arrest epileptic seizure	37164909	-	SNOMED
Automatisms	Neonatal focal automatism epileptic seizure	1244422	-	SNOMED
Automatisms	Focal onset automatism epileptic seizure	37167413	-	SNOMED
Diagnostics tools				
EEG	Electroencephalogram	4181917	-	SNOMED
EEG	EEG finding	4104661	-	SNOMED

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
EEG	Ambulatory EEG	4098779	-	SNOMED
Video EEG	Video EEG	4102081	-	SNOMED
Video EEG	Video EEG study	37020959	-	LOINC
Pre-specified conditions				
Hypoxic-ischaemic encephalopathy	Hypoxic ischaemic encephalopathy	45766189	-	SNOMED
Stroke or haemorrhage	Ischemic stroke	4310996	-	SNOMED
Stroke or haemorrhage	Haemorrhagic stroke	35609033	-	SNOMED
Stroke or haemorrhage	Neonatal stroke	4159152	-	SNOMED
Infections	Meningitis	435785	-	SNOMED
Infections	Sepsis	132797	-	SNOMED
Infections	Encephalitis caused by human herpes simplex virus	4322568	-	SNOMED
Cortical malformations	Congenital anomaly of brain	377085	-	SNOMED
Cortical malformations	Polymicrogyria	35623413	-	SNOMED
Cortical malformations	Lissencephaly	4070082	-	SNOMED
Errors of metabolism	Glucose transporter protein type 1 deficiency syndrome	40482875	-	SNOMED
Errors of metabolism	Pyridoxine-dependent epilepsy	42535907	-	SNOMED
Errors of metabolism	X-linked creatine deficiency	44782653	-	SNOMED
Errors of metabolism	Phenylketonuria	40388130	-	SNOMED
Errors of metabolism	Maple syrup urine disease	4100475	-	SNOMED
Genetic aetiologies	Self-limited familial neonatal epilepsy	4043411	-	SNOMED
Genetic aetiologies	Early infantile epileptic encephalopathy with suppression bursts	4043414	-	SNOMED
Genetic aetiologies	Early myoclonic encephalopathy	4186206	-	SNOMED
Outcomes				
Epilepsy	Epilepsy	380378	-	SNOMED
Cerebral palsy	Cerebral palsy	4134120	-	SNOMED
Development delay	Developmental delay	436077	-	SNOMED
Intellectual disability	Intellectual disability	40277917	-	SNOMED
Behavioural disorders	Attention deficit hyperactivity disorder	438409	-	SNOMED
Behavioural disorders	Autism spectrum disorder	439776	-	SNOMED

Table S2. Preliminary list of medicines definitions.

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
Phenobarbital	Phenobarbital	Antiepileptics	N03AA02	734275	Yes
Phenytoin	Phenytoin	Antiepileptics	N03AB02	740910	Yes
Fosphenytoin	Fosphenytoin	Antiepileptics	N03AB05	713192	Yes
Levetiracetam	Levetiracetam	Antiepileptics	N03AX14	711584	Yes
Midazolam	Midazolam	Hypnotics and sedatives	N05CD08	708298	Yes
Lidocaine	Lidocaine	Anaesthetics	N01BB02	989878	Yes
Carbamazepine	Carbamazepine	Antiepileptics	N03AF01	740275	Yes
Pyridoxine	Pyridoxine	Vitamin preparations	A11HA02	19005046	Yes

ANNEX V: ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Neonatal seizures: Incidence, prevalence, patient characterisation, and treatments in European countries

EU PAS Register® number: Study not registered yet
Study reference number (if applicable): P4-C1-014

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]	☒	☐	☐	5
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)	☒	☐	☐	7
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?	☒	☐	☐	8.4
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)	☐	☐	☒	

Comments:

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

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)	☒	☐	☐	8.6.1
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)	☒	☐	☐	8.8
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)	☒	☐	☐	ANNEX I
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	ANNEX I
9.1.3 Covariates and other characteristics?	☒	☐	☐	ANNEX I
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)	☒	☐	☐	ANNEX I
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	ANNEX I
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	ANNEX I
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	ANNEX I
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	ANNEX I
9.3.3 Covariates and other characteristics?	☒	☐	☐	ANNEX I
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☒	☐	☐	ANNEX I

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?	☒	☐	☐	8.8
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?	☐	☐	☒	

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)	☒	☐	☐	ANNEX III
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)	☒	☐	☐	ANNEX II

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?	☒	☐	☐	ANNEX II
13.3 Have data protection requirements been described?	☒	☐	☐	ANNEX III

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?	☒	☐	☐	4

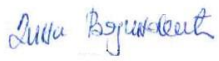
Comments:

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?	☒	☐	☐	ANNEX III

Comments:

Name of the main author of the protocol: Dina Vojinovic

Date: 05/12/2025

Signature: 

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.