Study Protocol

P4-C2-012

DARWIN EU® - RR Childhood hypertension and sartans prescribing in children

18/08/2025

Version 2.0

Authors: Ellen Gerritsen, Dina Vojinovic

Public

CONTENTS

Study title[1]	DARWIN EU® - RR Childhood hypertension and sartans prescribing in children
Protocol version	V2.0
Date	18/08/2025
EUPAS number	EUPAS1000000714
Active substance	Sartans drugs classes with corresponding WHO ATC code (classified at 4th level): • C09CA: Angiotensin II receptor blockers, plain • C09DA: Angiotensin II receptor blockers and diuretics • C09DB: Angiotensin II receptor blockers and calcium channel blockers • C09DX: Angiotensin II receptor blockers, other combinations Antihypertensive drug classes with corresponding WHO ATC code (classified at 2nd level): • C03: Diuretics • C07: Beta blocking agents • C08: Calcium channel blockers • C09: Agents acting on the renin-angiotensin system
Medicinal product	n/a
Research question and objectives	Research question: What is the real-world prevalence of childhood hypertension and antihypertensive medication prescribing among patients with childhood hypertension over time across Europe? Study objectives: 1. To estimate the annual prevalence of childhood hypertension in the paediatric population. Results will be stratified by age group (children vs. adolescents), sex, and type of hypertension (primary vs. secondary). 2. To estimate the annual prevalence of prescribing of sartans and other antihypertensive medications in patients with childhood hypertension. Results will be stratified by drug class, age group (children vs. adolescents), sex, and type of hypertension (primary vs. secondary).
Countries of study	Finland, Germany, Hungary, Norway, Spain
Authors	Ellen Gerritsen, e.gerritsen@darwin-eu.org Dina Vojinovic, d.vojinovic@darwin-eu.org

[1]This is a routinely repeated study from P4-C1-015 with EUPAS EUPAS1000000714.

1. LIST OF ABBREVIATIONS

Acronyms/term	Description
ATC	Anatomical Therapeutic Chemical
BIFAP	Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público
CDM	Common Data Model
CC	Coordinating centre
CHT	Childhood hypertension
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DQD	Data Quality Dashboard
DOI	Declaration of Interests
DQD	Data Quality Dashboard
DRE	Digital Research Environment
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 TaUH-Pirha	Tampere University Hospital patient cohort
GDPR	General Data Protection Regulation
ICD	International Classification of Diseases
InGef RDB	InGef Research Database
IP	Inpatient
IRB	Institutional Review Board
NLHR	Norwegian Linked Health Registry data
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
RxNorm	Medical prescription normalized
SIDIAP	The Information System for Research on Primary Care
SNOMED	Systematized Nomenclature of Medicine
SUCD	Semmelweis University Clinical Data
WHO	World Health Organisation

2. TITLE

DARWIN EU® - RR Childhood hypertension and sartans prescribing in children

3. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Ellen Gerritsen Dina Vojinovic	IQVIA
Data Scientist	Akram Mendez Isabella Kaczmarczyk	IQVIA
Study Manager	Natasha Yefimenko	Erasmus MC
Data Partner*	**Names**	**Organisation**
FinOMOP-TaUH Pirha	Sampo Kukkurainen Leena Hakkarainen Kati Kristiansson	FinOMOP Tampere
InGef RDB	Annika Vivirito Josephine Jacob Raeleesha Norris Alexander Harms	InGef - Institut für angewandte Gesundheitsforschung Berlin GmbH
SUCD	Loretta Kiss Ágota Mészáros Tibor Héja Zsolt Bagyura András Sallai	Semmelweis University
NLHR	Saeed Hayati Nhung Trinh Hedvig Marie Egeland Nordeng	Norwegian Institute of Public Health
BIFAP	Ana LIorente-Garcia Miguel Angel Macia Martinez María del Mar Martín Pérez Elvira Rubio Esparza Alicia Peñaranda Navazo	Agencia Española de Medicamentos y Productos Sanitarios
SIDIAP	Laura Granés González Agustina Giuliodori Picco Irene López Sánchez Anna Palomar Cros	Institute for Primary Health Care Research Jordi Gol i Gurina

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

4. ABSTRACT

Title

DARWIN EU® – RR Childhood hypertension and sartans prescribing in children

Rationale and background

Childhood hypertension (CHT), defined as elevated blood pressure in children and adolescents, is a significant health concern with implications for both short- and long-term health outcomes. CHT can be classified into two main categories. Primary hypertension refers to cases without an identifiable underlying cause, while hypertension that results from a specific underlying, potentially reversible cause, is classified as secondary hypertension. Among the pharmacological options available for managing CHT, angiotensin receptor blockers, commonly referred to as sartans, are among the recommend first-line antihypertensive treatments. However, real-world data on prevalence of CHT and the prescribing patterns of sartans and other antihypertensive medications in the paediatric populations remain limited. This study aims to generate real-world evidence on the prevalence of CHT and the prescribing patterns of sartans and other antihypertensive medication among individuals with CHT across Europe to support regulatory decision-making and inform clinical practice.

Research question and objectives

Research question

What is the real-world prevalence of childhood hypertension and antihypertensive medication prescribing among patients with childhood hypertension over time across Europe?

Study objectives

1. To estimate the annual prevalence of childhood hypertension in the paediatric population. Results will be stratified by age group (children vs. adolescents), sex, and type of hypertension (primary vs. secondary).
2. To estimate the annual prevalence of prescribing of sartans and other antihypertensive medications in patients with childhood hypertension. Results will be stratified by drug class, age group (children vs. adolescents), sex, and type of hypertension (primary vs. secondary).

Methods

Study design

- Descriptive disease epidemiology study employing a population-level cohort to estimate the prevalence of childhood hypertension in the paediatric population (*objective 1*)
- Drug utilisation study employing a population-level cohort to estimate the prevalence of prescribing of sartans and other antihypertensive medication in individuals with childhood hypertension (*objective 2*)

The study period for recruitment is from 1st of January 2015 to 31st of December 2024 (or latest date available).

  - Index date (*objective 1*): The earliest date within the study period on which an individual aged 18 years or younger was recorded in the data source.
  - Index date (*objective 2*): The earliest date within the study period on which an individual has a recorded diagnosis of CHT.

Individuals are followed up until 1) end of study period (31st of December 2024), 2) end of data availability, 3) loss to follow up, 4) age ≥ 19 years, or 5) death, whichever came first.

Population

The study population includes:

- All individuals aged 18 years and younger who are registered in the database during the recruitment period (from 01/01/2015 to 31/12/2024 (or latest date available)) *(objective 1*).
- All individuals aged 18 years or younger with a recorded diagnosis of childhood hypertension in the database during the study period (*objective 2*).

Variables

Outcomes

Condition of interest: childhood hypertension (CHT)

Drugs of interest: sartans (WHO ATC codes C09CA, C09DA, C09DB, and C09DX) and other antihypertensive medication drug classes (WHO ATC codes C03, C07, C08, and C09)

Relevant covariates: age group (children aged >0 to <13 years vs. adolescents aged ≥13 to <19 years), sex, and type of hypertension (primary vs. secondary)

Data source

1. Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)
2. Germany: InGef Research Database (InGef RDB)
3. Hungary: Semmelweis University Clinical Data (SUCD)
4. Norway: Norwegian Linked Health Registry data (NLHR)
5. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
6. Spain: The Information System for Research on Primary Care (SIDIAP)

Study size

No sample size has been calculated, as this is an exploratory study which will not test a specific hypothesis. Based on a preliminary feasibility assessment, the estimated number of record counts for CHT in the databases included in this study ranges from 5,800 (FinOMOP-TaUH Pirha) to 30,400 (SUCD). The estimated number of record counts for sartans in children in the databases included in this study ranges from 1,100 (SUCD) to 63,600 (BIFAP).

Statistical analysis

Annual period prevalence (expressed as proportion) of 1) CHT among paediatric population and 2) pre-specified antihypertensive medication among individuals with CHT will be estimated. Prevalence will be calculated overall for children aged ≤18 years old, and stratified by age categories, sex, and type of hypertension.

The statistical analyses will be conducted on OMOP CDM mapped data using the *IncidencePrevalence* R package.

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

5. AMENDMENTS AND UPDATES

None.

6. MILESTONES

Study milestones and deliverables	Planned dates
Final Study Protocol	August 2025
Creation of Analytical code	August/September 2025
Execution of Analytical Code on the data	September/October 2025
Deadline DARWIN EU® CC receives results from Data Partners	1st of December 2025
Draft Study Report	19th of December
Revision of Study Report	December 2025/January 2026
Final Study Report	30th January 2026

7. RATIONALE AND BACKGROUND

Childhood hypertension (CHT), defined as elevated blood pressure in children and adolescents, is a significant health concern due to its association with organ damage during childhood, increased risk of hypertension as a young adult, and serious adverse cardiovascular outcomes in adulthood.(1-3) CHT can be classified into two main categories. Primary hypertension refers to cases without an identifiable underlying cause, while hypertension that results from a specific underlying, potentially reversible cause, is classified as secondary hypertension.(4, 5) Secondary hypertension is frequently caused by coarctation of the aorta or renal diseases, but can also be triggered by other causes.(4) Among the pharmacological options available for managing CHT, angiotensin receptor blockers, commonly referred to as sartans, are among the recommend first-line antihypertensive treatments.(6, 7) However, real-world data on prevalence of CHT and the prescribing patterns of sartans and other antihypertensive medications in the paediatric populations remain limited. This study aims to generate real-world evidence on the epidemiology of CHT and prescribing patterns of sartans and other antihypertensive medication among individuals with CHT across Europe to support regulatory decision-making and inform clinical practice.

This study is a routine repeated study of a previous DARWIN EU® study P4-C1-015, which focused on CHT and sartans prescribing in children. This current study is now being repeated to include a broader network of data sources within the DARWIN EU® initiative.

8. RESEARCH QUESTION AND OBJECTIVES

Research questions

What is the real-world prevalence of childhood hypertension and antihypertensive medication prescribing among patients with childhood hypertension over time across Europe?

Research objectives

The aim of this study is to assess the prevalence of childhood hypertension and of sartans and other antihypertensive medication prescribing among patients with childhood hypertension in European countries.

The specific objectives of this study are:

1. To estimate the annual prevalence of childhood hypertension (CHT) in the paediatric population. Results will be stratified by age group (children vs. adolescents), sex, and type of hypertension (primary vs. secondary).
2. To estimate the annual prevalence of sartans and other antihypertensive medication prescribing in patients with childhood hypertension (CHT). Results will be stratified by drug class, age group (children vs. adolescents), sex, and type of hypertension (primary vs. secondary).

9. RESEARCH METHODS

9.1. Study design

A cohort study will be conducted using routinely collected health data from 7 databases from 5 countries across Europe and in 5 EU member states. The study will comprise of:

- A descriptive disease epidemiology study will be conducted to address *objective 1*, assessing the prevalence of CHT in the paediatric population.
- A drug utilisation study will be conducted to address *objective 2*, assessing the prevalence of sartans and other antihypertensive prescribing among individuals diagnosed with CHT during the study period.

Figure 1 provides an overview of the study design by depicting when inclusion criteria, exclusion criteria, and covariates will be assessed respective to the cohort entry date.

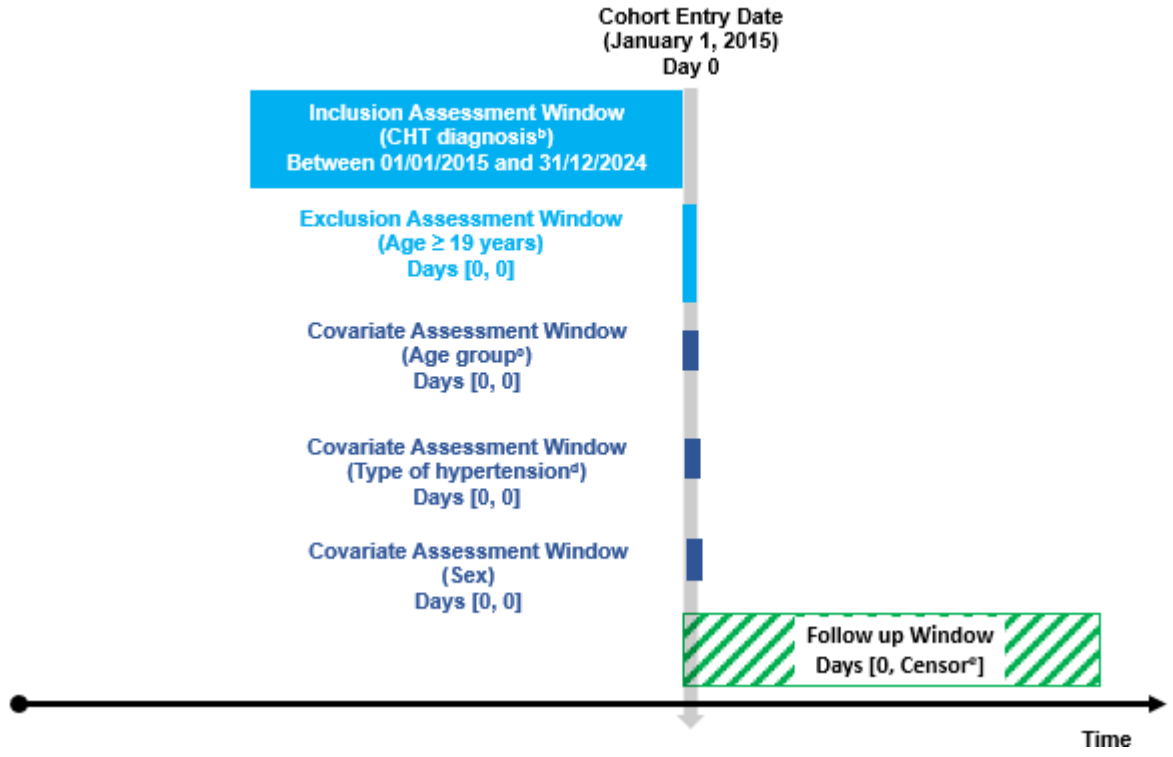

Figure 1. Graphical depiction of the study design.

a. Prevalence of sartans will be assessed per drug class at WHO ATC level 4 (WHO ATC codes C09CA and C09DA-DX) and of other pre-specified antihypertensive medication per drug class at WHO ATC level 2 (WHO ATC codes C03, C07, C08, C09)
b. The inclusion criterion of CHT diagnosis only applies for *objective 2*.
c. Stratification into 1) children aged between >0 and <13 years and 2) adolescents aged between ≥13 and <19 years
d. Stratification into 1) primary and 2) secondary hypertension
e. Earliest of 1) death, 2) disenrollment, 3) end of the study period, and 4) age ≥ 19 years

CHT = childhood hypertension

9.2. Study setting and data sources

This study will be conducted using routinely collected data from 1 registry, 2 hospital care, 2 primary care and hospital care, and 1 claims data sources in the DARWIN EU® network of data partners from 5 European countries in 5 EU member states. All data were a priori mapped to the OMOP CDM.

Data sources

1. Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)
2. Germany: InGef Research Database (InGef RDB)
3. Hungary: Semmelweis University Clinical Data (SUCD)
4. Norway: Norwegian Linked Health Registry data (NLHR)
5. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
6. Spain: The Information System for Research on Primary Care (SIDIAP)

Data Selection

These data sources fulfil the criteria required in terms of data quality, completeness, timeliness, and representativeness for population-level descriptive epidemiology and patient-level drug utilisation studies while covering different regions of Europe.

When it comes to assessing the reliability of data sources, the data partners are asked to describe their internal data quality process on the source data as part of the DARWIN EU® onboarding procedure. To further ensure data quality, we utilised the *Achilles* tool,(8) which systematically characterises the data and generates data characteristics such as age distribution, condition prevalence per year, and data density. Data density includes information on 1) monthly record counts by data domain (which offers insights into data collection patterns and the start date of each data source) and 2) measurement value distribution (i.e., min, max, quartiles for numeric values per measurement concept and per unit and counts for discrete measurement-value pairs). The latter can be compared against expectations for the data based on predefined standards, historical trends, or known epidemiological patterns to identify potential anomalies or inconsistencies. Additionally, the data quality dashboard (DQD) provides more objective checks (see Section D1.3.5.2 on Complete Data Quality Assurance Package) on plausibility of data completeness, consistency, and conformity across the data sources.

In terms of relevance, the selection of data sources was based on the availability of data on CHT and prescribing of sartans and other antihypertensive medication to perform the described analyses. In addition, the data sources were chosen considering their ability to support timely IRB approvals, thus ensuring alignment with the timeline established by stakeholders for the conduct of this study.

The DARWIN EU® portal, as well as information from the onboarding documents, were used to assess whether data sources have information on CHT and prescribing of sartans and other antihypertensive medication. Data within the DARWIN EU® portal is maintained up to date by extracting the release dates for each dataset in the network and monitoring when data are out-of-date with the expected refresh cycle (typically quarterly or half-yearly). In addition, it is important to have clear understanding of the time covered by each released data source, as this can vary across different domains. To facilitate this, the *CDMOnboarding* (and *Achilles*) packages (8) contain a 'data density' plot. This plot displays the number of records per OMOP domain monthly. This allows to get insights when data collection started, when new sources of data were added, and until when data was included. In addition, at time of inviting data partners, they were informed about study objectives and asked whether they could participate in the study.

More general-purpose diagnostic tools, *CohortDiagnostics* (9) and *DrugExposureDiagnostics* (10), have been developed. The *CohortDiagnostics* package provides additional insights into cohort characteristics,

record counts, and index event misclassification. The *DrugExposureDiagnostics* package evaluates ingredient-specific attributes and patterns in drug exposure records. Upon finalisation of the study protocol and creation of the disease and drug cohorts of interest by DARWIN EU® Coordination Centre, these packages will be executed in each data sources by each data partners.

Data source justification and key characteristics

General information on the data sources planned to include in this study is provided in **ANNEX I. Data sources description**. The key characteristics are described below per data source.

Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha), Finland

FinOMOP-TaUH Pirha will be included in this study, as it is a hospital data source that provides relevant information on CHT and the prescribing of sartans and other antihypertensive medication among individuals with CHT in the general paediatric population (children <19 years).

Based on a preliminary feasibility assessment, the expected number of person counts for CHT is approximately 5,800 and the number of sartans prescriptions in children is estimated at 1,200.

Data availability and follow-up in FinOMOP-TaUH Pirha are sufficient to support the study objectives. FinOMOP-TaUH Pirha has been collecting data since 2007, with the most recent data extraction date 03/2025. This aligns with the study period. The median follow-up of the first observation period in FinOMOP-TaUH Pirha is 4,230 days (IQR: 374–7,979 days).

There are no study specific limitations associated with FinOMOP-TaUH Pirha.

Finally, IRB approval for FinOMOP-TaUH Pirha is estimated to take 1 week, which facilitates the timely execution of this study within the current study timelines.

InGef Research Database (InGef RDB), Germany

InGef RDB will be included in this study, as it is a claims data source that provides relevant information on CHT and the prescribing of sartans and other antihypertensive medication among individuals with CHT in the general paediatric population (children <19 years).

Based on a preliminary feasibility assessment, the expected number of person counts for CHT is approximately 13,300 and the number of sartans prescriptions in children is estimated at 20,300. Data availability and follow-up in InGef RDB are sufficient to support the study objectives.

Data availability in InGef RDB starts in 2015 and the date of most recent data extraction is 12/2024. This aligns with the study period and the median follow-up of the first observation period in InGef RDB is 3,560 days (IQR: 1,401–3,652 days).

There are some potential study specific limitations associated with InGef RDB. InGef RDB outpatient data is dated to the end of every quarter, i.e., all observations between January 1st and March 31st are dated on March 31st. This will result in potential misclassification in diagnosis and treatment, where date of treatment might fall prior to the date of CHT diagnosis as recorded in the database. To account for this, the date of CHT diagnosis will be moved to the beginning of the quarter, therefore ensuring that medication use will start after CHT diagnosis. This approach could potentially lead to a small-time increase (maximum 3 months) for the time contributed by the CHT patients towards the corresponding denominator cohort.

Nevertheless, this will likely not greatly influence the final results.

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

Semmelweis University Clinical Data (SUCD), Hungary

SUCD will be included in this study, as it is a hospital data source that provides relevant information on CHT and the prescribing of sartans and other antihypertensive medication among individuals with CHT in the general paediatric population (children <19 years).

Based on a preliminary feasibility assessment, the expected number of person counts for CHT is approximately 30,400 and the number of sartans prescriptions in children is estimated at 1,100 in SUCD.

Data availability and follow-up in SUCD are sufficient to support the study objectives. SUCD has been collecting data since 2005, and the date of most recent data extraction is 11/2024. This aligns with the study period. The median follow-up of the first observation period in SUCD is 266 days (IQR: 0–2,165 days).

There are no study specific limitations associated with SUCD. The study period for SUCD will be January 2015 until November 2024.

Finally, IRB approval for SUCD is estimated to take 3 months, which facilitates the timely execution of this study within the current study timelines.

Norwegian Linked Health Registry data (NLHR), Norway

NLHR will be included in this study, as it is a registry data source that provides relevant information on CHT and the prescribing of sartans and other antihypertensive medication among individuals with CHT in the general paediatric population (children <19 years).

Based on a preliminary feasibility assessment, the expected number of person counts for CHT is approximately 29,800 and the number of sartans prescriptions in children is estimated at 12,800 in NLHR.

Data availability and follow-up in NLHR is sufficient to support the study objectives. NLHR has been collecting data since 2008, with the most recent data extraction dated 12/2023. This aligns with the study period. The median follow-up of the first observation period in NLHR is 5,843 days (IQR: 0–5,843 days).

There are no study specific limitations associated with NLHR. The study period for NLHR will be from January 2019 until December 2023, as reliable drug description data is available from January 2019 onwards.

Finally, IRB approval for NLHR is estimated to take 1 month, which facilitates the timely execution of this study within the current study timelines.

Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP), Spain

BIFAP will be included in this study, as it is a primary care and hospital data source that provides relevant information on CHT and the prescribing of sartans and other antihypertensive medication among individuals with CHT in the general paediatric population (children <19 years).

Based on a preliminary feasibility assessment, the expected number of person counts for CHT is approximately 11,000 and the number of sartans prescriptions in children is estimated at 63,600 in BIFAP.

Data availability and follow-up in BIFAP is sufficient to support the study objectives. BIFAP has been collecting data since 2001, with the most recent data extraction dated 12/2024. This aligns with the study period. The median follow-up of the first observation period in BIFAP is 4,016 days (IQR: 1,811–6,263 days).

There are no study specific limitations associated with BIFAP.

Finally, IRB approval for BIFAP is estimated to take 2 months, which facilitates the timely execution of this study within the current study timelines.

The Information System for Research on Primary Care (SIDIAP), Spain

SIDIAP will be included in this study, as it is a primary care data source that provides relevant information on CHT and the prescribing of sartans and other antihypertensive medication among individuals with CHT in the general paediatric population (children <19 years).

Based on a preliminary feasibility assessment, the expected number of person counts for CHT is approximately 9,500 and the number of sartans prescriptions in children is estimated at 19,800 in SIDIAP.

Data availability and follow-up in SIDIAP is sufficient to support the study objectives. SIDIAP has been collecting data since 2006, with the most recent data extraction dated 06/2023. This aligns with the study period. The median follow-up of the first observation period in SIDIAP is 5,670 days (IQR: 2,223–6,389 days).

There are no study specific limitations associated with SIDIAP. The study period for SIDIAP will be January 2015 until June 2023.

Finally, IRB approval for SIDIAP is estimated to take 2 months, which facilitates the timely execution of this study within the current study timelines.

9.3. Study period

The study period is from 01/01/2015 to 31/12/2024 or the most recent data available for each contributing data source.

It should be noted for several data sources, the availability of the accurate data deviates from the start or end date of the study period. Detailed information about the study period per data partner can be found in **Section 9.2**.

9.4. Follow-up

For the descriptive disease epidemiology study (*objective 1*), follow-up will start on the earliest date within the study period (01/01/2015 – 31/12/2024) when an individual is recorded in the data source while aged ≤18 years.

For the drug utilisation study (*objective 2*), follow-up will start on the earliest date on which an individual has a record of CHT diagnosis while aged ≤18 years.

End of follow-up will be defined as the earliest of 1) loss to follow-up, 2) death, 3) end of observation period (the latest available data), or 4) aged ≥19 years or older, whichever occurs first.

Estimating prevalence requires an appropriate denominator population and the corresponding observation time. Study participants will begin contributing person-time at risk as described above in **Section 9.4 Follow-up**.

An illustrative example of entry and exit into the denominator population is shown in **Figure 2**. In this example, the observation period of person IDs 1 and 2 starts before the study start date and the observation period ends after the study end date, so this person will contribute during the complete study period. Person ID 3 leaves when exiting the data source (the end of the observation period). Person ID 4 enters the study when their observation period starts. Lastly, person ID 5 has two observation periods in the data source. The first period contributes time from the study start until the end of the observation period, the second starts contributing time again once the observation period starts and exits at study end date.

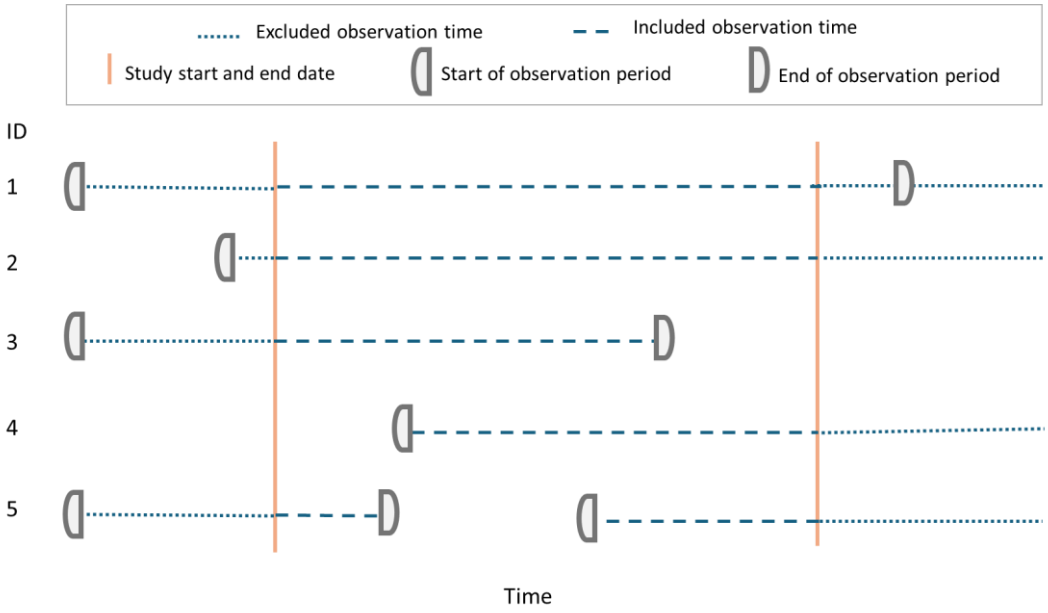

Figure 2. Included observation time for the denominator population.

9.5. Study population with inclusion and exclusion criteria

For prevalence calculations of CHT (*objective 1*), the study population will include all individuals who are 18 years or younger and registered in the data source between the 1st of January 2015 and 31st of December 2024 (or the latest data available of the respective data source).

For prevalence calculations of prescribing of sartans and other antihypertensive medications in patients with CHT (*objective 2*), the study population will include all individuals registered in the data source with a condition occurrence of CHT, defined as a SNOMED diagnostic code for hypertension, in individuals who are 18 years and younger, between the 1st of January 2015 and 31st of December 2024 (or the latest data available of the respective data source). Only individuals who are 18 years or younger at the date of prescription (index date) will be included. The preliminary concept sets used for the identification of individuals with CHT are described in **ANNEX III.**

9.6. Variables

9.6.1. Exposure

Not applicable.

9.6.2. Outcome

Objective 1:

The outcome for this objective is as follows:

- Occurrence of CHT, defined as a recorded SNOMED diagnostic code for hypertension in individuals aged 18 years or younger.

Objective 2:

The outcome for this objective is as follows:

- Prescribing of pre-specified antihypertensive medication among individuals with CHT, defined as a recorded RxNorm prescription of pre-specified antihypertensive medication in individuals diagnosed with hypertension and aged 18 years or younger.

- Sartans (WHO ATC codes C09CA, C09DA, C09DB, C09DX) will be assessed per drug class at WHO ATC level 4.
- Other pre-specified antihypertensive medication (WHO ATC codes C03, C07, C08, C09) will be assessed per drug class at WHO ATC level 2.

The preliminary concept sets used for the identification of outcomes are described in **ANNEX III**. These codes will be refined during the study execution following the DARWIN EU® phenotyping standard processes, which involves the review of phenotypes by the study team and EMA.

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

All objectives:

The covariates for these objectives are as follows and will be assessed at the index date corresponding to each objective:

- Age groups defined at index date namely
  - Overall paediatric population: individuals aged between >0 and <19 years
  - Children: individuals aged between >0 and <13 years
  - Adolescents: individuals aged between ≥13 and <19 years
- Sex
  - Overall
  - Male
  - Female
- Type of hypertension namely
  - Overall
  - Primary hypertension
  - Secondary hypertension

The preliminary concept sets used for the identification of the type of hypertension are described in **ANNEX III**. These codes will be refined during the study execution following the DARWIN EU® phenotyping standard processes, which involves the review of phenotypes by the study team and EMA.

9.7. Study size

No sample size has been calculated, as this is an exploratory study which will not test a specific hypothesis. Based on a preliminary feasibility assessment, the estimated number of record counts for CHT in the databases included in this study range from 5,800 (FinOMOP-TaUH Pirha) to 30,400 (SUCD). The estimated number of record counts for sartans in children in the databases included in this study range from 1,100 (SUCD) to 63,600 (BIFAP).

9.8. Analysis

9.8.1. 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 patient-level 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 II.**

Additional information section Quality Control), the final package will be released in a version-controlled study repository for execution against all the participating data sources.

The data partners will locally execute the analytics against the OMOP CDM in R Studio and review and approve the default aggregated results. They will then be made available to the Principal Investigators and study team in secure online repository (Data Transfer Zone). All results will be locked and timestamped for reproducibility and transparency. The study results of all data sources are checked after which they are made available to the team and the Study Dissemination Phase can start. All results are locked and timestamped for reproducibility and transparency.

9.8.2. Patient privacy protection

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.

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

R-packages

The prevalence of CHT in the paediatric population and prescriptions of sartans and other antihypertensive medication among individuals with CHT will be calculated based on OMOP CDM mapped data using the *IncidencePrevalence* R package, developed by DARWIN EU® (https://github.com/darwin-eu/IncidencePrevalence).

Prevalence of CHT (*objective 1*) and of sartans and other antihypertensive medication prescriptions among individuals with CHT (*objective 2*)

Prevalence will be calculated as annual period prevalence, which summarises the total number of individuals who are diagnosed with childhood hypertension (*objective 1*) or the number of individuals with childhood hypertension who use sartans or other pre-specified antihypertensive medication (*objective 2*) during a given year divided by the population at risk of getting exposed during that year. Therefore, period prevalence gives the proportion of individuals exposed at any time during a specified interval. Binomial 95% confidence intervals will be calculated.

An illustration of the calculation of period prevalence is shown below in **Figure 3Error! Reference source not found.**. Between time t+2 and t+3, two of the five study participants are users of pre-selected drug of interest 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 being a user of pre-selected drug of interest.

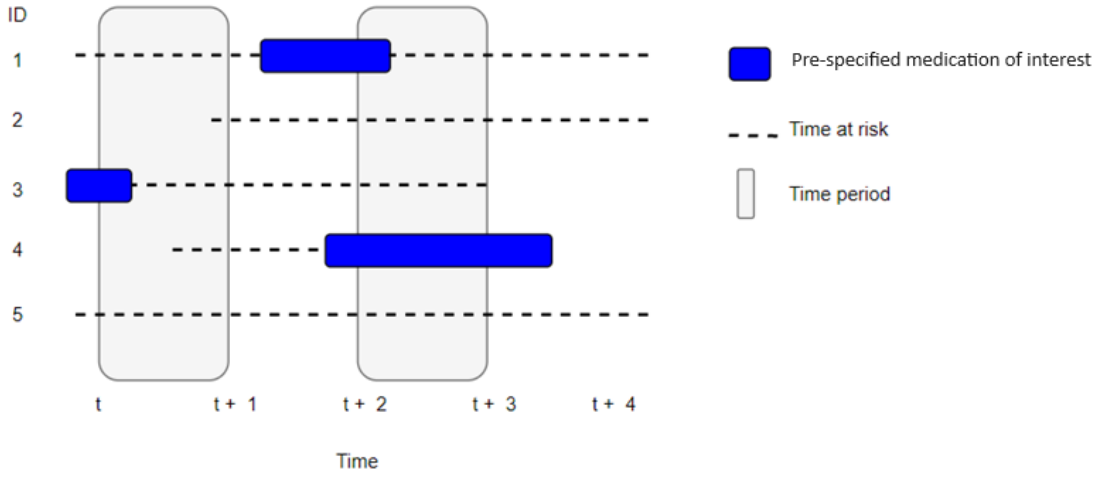

Figure 3. Period prevalence example.

Prevalence estimates will be stratified by type of hypertension, sex, and age group. The following types of hypertension will be used for stratification: primary hypertension and secondary hypertension. The following age groups will be used for stratification: children (aged between >0 and <13 years) and adolescents (aged between ≥13 and <19 years). Age at index date will be calculated using January 1st of the year of birth as proxy for the actual birthday. Date/month is either not present or cannot be made available for governance reasons. If available, date is often set to first of the month for patient's privacy.

Methods to deal with missing data

For the drug utilisation studies we assume that the absence of a prescription record means that the person does not receive the respective drug. For indications, we assume that the missingness of a record of the respective condition mean that that condition is not the indication for the drug prescription.

9.8.4. Output

Output will include the following:

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

- Table 1. Attrition table *(objective 1-2).*
- Figure 1. Annual prevalence of CHT in the overall paediatric population per database (*objective 1*).
- Figure 2. Annual prevalence of CHT in the paediatric population stratified by age group per database (*objective 1*).
- Figure 3. Annual prevalence of CHT in the paediatric population stratified by sex per database (*objective 1*).
- Figure 4. Annual prevalence of CHT in the paediatric population stratified by type of hypertension per database (*objective 1*).
- Figure 5. Annual prevalence of sartans and other pre-specified antihypertensive medication prescriptions in individuals with CHT per database (*objective 2*).
- Figure 6. Annual prevalence of sartans and other pre-specified antihypertensive medication prescriptions in individuals with CHT stratified by age group per database (*objective 2*).
- Figure 7. Annual prevalence of sartans and other pre-specified antihypertensive medication prescriptions in individuals with CHT stratified by sex per database (*objective 2*).

- Figure 8. Annual prevalence of sartans and other pre-specified antihypertensive medication prescriptions in individuals with CHT stratified by type of hypertension per database (*objective 2*).

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

9.9. Evidence synthesis

Results from analyses described in **Section 9.8** will be presented separately for each data source. No meta-analysis of results will be conducted.

10. STRENGTHS AND LIMITATIONS

The study will be informed by routinely collected health care data, and it is important to consider several factors that may influence the interpretation of the results. This study will include data from multiple healthcare settings and types of data from across 5 different European countries (Finland, Germany, Hungary, Norway, and Spain), including 1 registry, 2 hospital care, 2 primary care and hospital care, and 1 claims data source, to ensure a diverse sample. However, the results derived from these databases may not be generalisable to populations outside these countries or to other healthcare systems.

The denominator used to calculate prevalence will vary across data sources. Hospital-based datasets will include only paediatric population who had hospital encounters, while primary care and claims databases capture broader populations. These differences may affect the comparability of prevalence estimates across data sources. Therefore, prevalence estimates will be reported by data source type.

Electronic health records and claims data were collected for clinical or administrative purposes rather than primarily for research use. As a result, data may be incomplete. Additionally, recorded prescription does not necessarily indicate that the patient actually took the drug. Therefore, assumptions of actual use are made.

Differences in diagnostic criteria of CHT, coding practices, and the number of blood pressure measurements required may influence the identification of cases across data sources and affect prevalence estimates. Importantly, there is no universal classification of CHT, i.e., whether the diagnosis should be based on one or multiple blood pressure checks. This might affect the number of individuals diagnosed with CHT.

11. REFERENCES

1. Theodore RF, Broadbent J, Nagin D, Ambler A, Hogan S, Ramrakha S, et al. Childhood to Early-Midlife Systolic Blood Pressure Trajectories. Hypertension. 2015;66(6):1108-15.

2. Urbina EM, Mendizábal B, Becker RC, Daniels SR, Falkner BE, Hamdani G, et al. Association of Blood Pressure Level With Left Ventricular Mass in Adolescents. Hypertension. 2019;74(3):590-6.

3. Yang L, Magnussen CG, Yang L, Bovet P, Xi B. Elevated Blood Pressure in Childhood or Adolescence and Cardiovascular Outcomes in Adulthood. Hypertension. 2020;75(4):948-55.

4. Bassareo PP, Calcaterra G, Sabatino J, Oreto L, Ciliberti P, Perrone M, et al. Primary and secondary paediatric hypertension. Journal of Cardiovascular Medicine. 2023;24(Supplement 1):e77-e85.

5. Charles L, Triscott J, Dobbs B. Secondary Hypertension: Discovering the Underlying Cause. Am Fam Physician. 2017;96(7):453-61.

6. de Simone G, Mancusi C, Hanssen H, Genovesi S, Lurbe E, Parati G, et al. Hypertension in children and adolescents: A consensus document from ESC Council on Hypertension, European Association of Preventive Cardiology, European Association of Cardiovascular Imaging, Association of Cardiovascular Nursing & Allied Professions, ESC Council for Cardiology Practice and Association for European Paediatric and Congenital Cardiology. European Heart Journal. 2022;43(35):3290-301.

7. Siddiqi N, Shatat IF. Antihypertensive agents: a long way to safe drug prescribing in children. Pediatric Nephrology. 2020;35(11):2049-65.

8. DeFalco F, Ryan P, Schuemie M, Huser V, Knoll C, Londhe A, et al. Achilles: Achilles Data Source Characterization. R package version 1.7.2. 2023.

9. Gilbert J, Rao G, Schuemie M, Ryan P, Weaver J. CohortDiagnostics: Diagnostics for OHDSI Cohorts. R package version 3.3.0, https://github.com/OHDSI/CohortDiagnostics, https://ohdsi.github.io/CohortDiagnostics. 2024.

10. Inberg G, Burn E, Burkard T. DrugExposureDiagnostics: Diagnostics for OMOP Common Data Model Drug Records. R package version 1.0.9, https://github.com/darwin-eu/DrugExposureDiagnostics, https://darwin-eu.github.io/DrugExposureDiagnostics/. 2024.

11. Palmaro A, Gauthier M, Conte C, Grosclaude P, Despas F, Lapeyre-Mestre M. Identifying multiple myeloma patients using data from the French health insurance databases: Validation using a cancer registry. Medicine (Baltimore). 2017;96(12):e6189.

12. Vikkula J, Uusi-Rauva K, Ranki T, Toppila I, Aalto-Setälä M, Pousar K, et al. Real-world evidence of multiple myeloma treated from 2013 to 2019 in the Hospital District of Helsinki and Uusimaa, Finland. Future Oncol. 2023;19(30):2029-43.

12. ANNEXES

ANNEX I. Data sources description

Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha), Finland

TaUH Research Database includes all specialities/all patient groups treated in the Tampere University Hospital, secondary and tertiary care given in the region including given clinical and pathology diagnoses, diagnostic and therapeutic procedures, laboratory findings, radiology and pathology reports, medication given in the hospital and electronic prescriptions, and continuous medical records, including discharge letters since 2007.

InGef Research Database (InGef RDB), Germany

The InGef database comprises anonymized longitudinal claims data of about 10 million individuals across more than 50 statutory health insurance providers (SHIs) throughout Germany. Data are longitudinally linked over a period of currently ten years. Patients can be traced across health care sectors. All patient-level and provider-level data in the InGef research database are anonymised to comply with German data protection regulations and German federal law. German SHI claims data available in the InGef database includes information on demographics (year of birth, gender, death date if applicable, region of residence on administrative district level); hospitalizations; outpatient services (diagnoses, treatments; specialities of physicians); dispensing of drugs; dispensing of remedies and aids; and sick leave and sickness allowance times. In addition, costs or cost estimates from SHI perspective are available for all important cost elements. All diagnoses in Germany are coded using the International Classification of Diseases, version 10 in the German Modification (ICD-10-GM). The persistence (membership over time) is rather high in the InGef database: During a time period of 5 years (2009 to 2013), 70.6% of insurance members survived and remained insured with the same SHI without any gap in their observational time. Persons leaving one of the participating SHIs and entering another participating SHI can be linked during yearly database consistency updates and are thus not lost over time. The InGef database is dynamic in nature, i.e. claims data are updated in an ongoing process and new SHIs may join or leave the database. By law, only the last 10 years of data are allowed to be used. At every new release this window shifts, dropping older data and adding new data.

Semmelweis University Clinical Data (SUCD), Hungary

Semmelweis University is the largest provider of health care services in Hungary. Most of the departments cater for the most serious cases and patients requiring complex treatment, thus making the university a national health care provider. The overwhelming majority of patient data originates from Hungary, mainly from central region of the country: Budapest and Pest County. The database contains approximately 2 million individual patients across all care settings of the University since 2011. The hospital information system (MedSolution) is an integrated IT system provides functional support for inpatient and outpatient care processes and serves as an integrated platform for different diagnostic areas, and in some specific area it supports the registration of medications. It supports all kinds of hospital work processes from admission to discharge. The outpatient module serves as a platform for the registration of activities related to care episode within the outpatient specialist care. During the care provision data related to health state of the patient, the diagnosis, the documentation of requested examinations and medical consultations, prescribed medication, final reports, and performed interventions are recorded. The functions of the inpatient module assist the care provision within the inpatient settings. It documents the health state of the patient at admission and during the hospital stay, along with the anamnesis, diagnosis, the performed examinations and interventions, hospital final reports, and provided medication in some are of care provision such as chemotherapy. Among other modules the diagnostic module registers the requested laboratory and imaging examinations and records the laboratory results.

Norwegian Linked Health Registry data (NLHR), Norway

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

Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP), Spain

BIFAP (http://www.bifap.org/index_EN.html) is a longitudinal population-based data source of medical patient records of the Spanish National Health Service (SNS) from 9 participating Regions throughout Spain out of the 17 Spanish Regions. Population currently included represents 36% of the total Spanish population. Spain has a SNS that provides universal access to health services through the Regional Healthcare Services. Primary care physicians (PCPs), both general practitioners and paediatricians, have a central role. They act as gatekeepers of the system and also exchange information with other levels of care to ensure the continuity of care. Most (98.9%) of the population is registered with a PCP and, in addition, most drug prescriptions are written at the primary care level. BIFAP includes a collection of databases linked at individual patient level. The main one is the Primary care Database given the central role of PCPs in the SNS. Linked, there are additional important structural databases like the medicines dispensed at community pharmacies and the patients' hospital diagnosis at discharge. 7 out of the 9 regions have linkage to hospital data. However, hospital data is available for different time periods for each region. From 2014 onwards, linkage to hospital data is available for >68% of patients. Linkage to SARS-CoV-2 diagnostics test and COVID-19 vaccination registries are also included. Additional databases are also linked for a subset of patients (hospital pharmacy, cause of death registry). BIFAP program is a non-profit program financed by the Spanish Agency of Medicines and Medical Devices (AEMPS), a government agency belonging to the Ministry of Health in collaboration with the regional health authorities.

The Information System for Research on Primary Care (SIDIAP), Spain

The Information System for Research in Primary Care (SIDIAP) is a clinical database of anonymized patient records in Catalonia, Spain. The Spanish public healthcare system covers more than 98% of the population, and more than two thirds of the Catalan population see their GP at least once a year. The computerisation of the primary care patient records of the Catalan Health Institute (CHI) was complete in 2005. SIDIAP was designed to provide a valid and reliable database of information from clinical records of patients registered in primary care centres for use in biomedical research. SIDIAP contains data of anonymized patients' healthcare records for nearly six million people (approximately 80% of the Catalan population) registered in 287 primary care practices throughout Catalonia since 2005. It includes data collected by health professionals during routine visits in primary care, including anthropometric measurements, clinical diagnoses (International Classification of Diseases 10th revision ICD-10), laboratory tests, prescribed and dispensed drugs, hospital referrals, demographic, and lifestyle information. It was previously shown that SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions. The high quality of these data has been previously documented, and SIDIAP has been successfully applied to epidemiological studies of key exposures and outcomes. Quality checks to identify duplicate patient IDs are performed centrally at each SIDIAP database update. Checks for logical values and data harmonisation are performed. For biochemistry data, consistency for measurements taken in different laboratories is assessed, and unit conversion is undertaken when needed.

ANNEX II. Additional information

DATA MANAGEMENT

Data management

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

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

Data storage and protection

For this study, participants from various EU member states will process personal data from individuals which 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. 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

General data source quality control

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

Study specific quality control

When defining cohorts for indications, a systematic search of possible codes for inclusion will be identified using *CodelistGenerator* R package (https://github.com/darwin-eu/CodelistGenerator). This software allows

the user to define a search strategy and using this will then query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, the *CohortDiagnostics* R package (https://github.com/OHDSI/CohortDiagnostics) 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 study code will be based on the R packages to estimate Prevalence using the OMOP common data model. This 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 III. List of stand-alone documents

Preliminary lists of conditions concepts definition are provided in the tables below:

Table S1. Preliminary list of concept definition for childhood hypertension.

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Hypertension	Hypertensive disorder	316866	None	SNOMED

Table S2. Preliminary list of concept definition for primary hypertension

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Primary hypertension	Essential hypertension	320128	None	SNOMED
Primary hypertension	Benign essential hypertension complicating pregnancy, childbirth and the puerperium with postnatal complication	4062811	None	SNOMED
Primary hypertension	Benign essential hypertension complicating pregnancy, childbirth and the puerperium - not delivered	314423	None	SNOMED
Primary hypertension	Benign essential hypertension complicating pregnancy, childbirth and the puerperium - delivered with postnatal complication	320456	None	SNOMED
Primary hypertension	Benign essential hypertension complicating pregnancy, childbirth and the puerperium - delivered	314103	None	SNOMED

Table S3. Preliminary list of concept definition for secondary hypertension

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Secondary hypertension	Secondary pulmonary hypertension	4339214	None	SNOMED
Secondary hypertension	Secondary ocular hypertension due to ocular trauma	37208896	None	SNOMED
Secondary hypertension	Secondary hypertension	319826	None	SNOMED
Secondary hypertension	Pulmonary venous hypertension due to compression of pulmonary great vein	43020840	None	SNOMED
Secondary hypertension	Pulmonary hypertension in systemic disorder	44783636	None	SNOMED
Secondary hypertension	Pulmonary hypertension due to pulmonary veno-occlusive disease	604306	None	SNOMED
Secondary hypertension	Pulmonary hypertension due to pulmonary disease with mixed restrictive and obstructive patterns	605200	None	SNOMED

Secondary hypertension	Pulmonary hypertension due to pulmonary capillary hemangiomatosis	604305	None	SNOMED
Secondary hypertension	Pulmonary hypertension due to lung disease and/or hypoxia	44783628	None	SNOMED
Secondary hypertension	Pulmonary hypertension due to left heart disease	43020910	None	SNOMED
Secondary hypertension	Pulmonary hypertension due to haematological disorder	44782564	None	SNOMED
Secondary hypertension	Pulmonary hypertension due to developmental abnormality	605202	None	SNOMED
Secondary hypertension	Pregnancy-induced hypertension	4167493	None	SNOMED
Secondary hypertension	Pregnancy induced hypertension with pulmonary oedema	44784483	None	SNOMED
Secondary hypertension	Portal hypertension due to cystic fibrosis	45771017	None	SNOMED
Secondary hypertension	Heritable pulmonary arterial hypertension due to BMPR2 mutation	44783619	None	SNOMED
Secondary hypertension	Heritable pulmonary arterial hypertension due to ALK1 or endoglin mutation	44783620	None	SNOMED
Secondary hypertension	Benign intracranial hypertension due to hypervitaminosis A	44782842	None	SNOMED
Secondary hypertension	Benign intracranial hypertension due to drug	44782841	None	SNOMED

Preliminary lists of concepts definitions for drug classes of antihypertensive medication are provided in the tables below:

Table S4. Preliminary list of concept definition for sartans (WHO ATC level 4)

Concept name	ATC code	Concept ID	Include descendants
Angiotensin II receptor blockers (ARBs), plain	C09CA	21601823	Yes
Angiotensin II receptor blockers (ARBs) and diuretics	C09DA	21601833	Yes
Angiotensin II receptor blockers (ARBs) and calcium channel blockers	C09DB	21601841	Yes
Angiotensin II receptor blockers (ARBs), other combinations	C09DX	21601845	Yes

Table S5. Preliminary list of concept definition for pre-specified antihypertensive medication (non-sartans (WHO ATC level 2))

Concept name	ATC code	Concept ID	Include descendants
DIURETICS	C03	21601461	Yes

BETA BLOCKING AGENTS	C07	21601664	Yes
CALCIUM CHANNEL BLOCKERS	C08	21601745	Yes
GENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM	C09	21601782	Yes

ANNEX IV: ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

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

Study title:
DARWIN EU® - Childhood hypertension and sartans prescribing in children

EU PAS Register® number: EUPAS1000000714
Study reference number (if applicable): P4-C2-012

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]	☒	☐	☐	6
	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)	☒	☐	☐	8
	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)	☒	☐	☐	9.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	9.2
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	9.8
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?	☒	☐	☐	9.2, 9.5
4.2	Is the planned study population defined in terms of:				
	4.2.1 Study time period	☒	☐	☐	9.3
	4.2.2 Age and sex	☒	☐	☐	9.6
	4.2.3 Country of origin	☒	☐	☐	9.2
	4.2.4 Disease/indication	☒	☐	☐	9.6
	4.2.5 Duration of follow-up	☒	☐	☐	9.4
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	9.5

Comments:

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☐	☐	☒	9.6.1
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?	☐	☐	☒	
5.4	Is intensity of exposure addressed? (e.g. dose, duration)	☐	☐	☒	
5.5	Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	
5.6	Is (are) (an) appropriate comparator(s) identified?	☐	☐	☒	

Comments:

Section 6: Outcome definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
6.1	Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	9.6.2
6.2	Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	9.6.2
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)	☐	☐	☒	

Section 9: Data sources		**Yes**	**No**	**N/A**	**Section Number**
	9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	9.2, 9.6
	9.1.3 Covariates and other characteristics?	☒	☐	☐	9.2, 9.6
9.2	Does the protocol describe the information available from the data source(s) on:				
	9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☐	☐	☒	
	9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	9.2, 9.6
	9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	9.2, 9.6
9.3	Is a coding system described for:				
	9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☐	☐	☒	
	9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	9.6
	9.3.3 Covariates and other characteristics?	☒	☐	☐	9.6
9.4	Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☒	

Comments:

Section 10: Analysis plan		**Yes**	**No**	**N/A**	**Section Number**
10.1	Are the statistical methods and the reason for their choice described?	☒	☐	☐	9.8
10.2	Is study size and/or statistical precision estimated?	☐	☐	☒	9.7
10.3	Are descriptive analyses included?	☒	☐	☐	9.8
10.4	Are stratified analyses included?	☒	☐	☐	9.8
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 II

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.2 Are methods of quality assurance described?	☒	☐	☐	Annex II
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?	☒	☐	☐	10
12.1.2 Information bias?	☒	☐	☐	
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)	☒	☐	☐	9.2, 9.7

Comments:

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	9.2
13.2 Has any outcome of an ethical review procedure been addressed?	☒	☐	☐	9.2
13.3 Have data protection requirements been described?	☒	☐	☐	Annex II

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?	☒	☐	☐	5

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 II

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.2 Are plans described for disseminating study results externally, including publication?	☐	☐	☒	

Comments:

Name of the main author of the protocol: Dina Vojinovic

Date: 18th July 2025

Signature: 

ANNEX V: Glossary

Aggregated Data

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

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® utilizes the OMOP CDM maintained by the OHDSI community .

Complex Studies (C3)

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

Coordination Centre (CC)

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

Data Access

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

Data Quality Framework

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

Data Source

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

DARWIN EU®

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

EMA (European Medicines Agency)

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

Evidence Generation

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

Federated Network

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

GDPR (General Data Protection Regulation)

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

Health Technology Assessment (HTA)

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

Metadata

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

Off-the-Shelf Studies (OTS)

Studies for which a standard protocol per study/analysis type and standardized 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 utilization 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 individual patients, often de-identified, used for longitudinal or detailed analyses.

OMOP (Observational Medical Outcomes Partnership)

A common data model (CDM) that standardizes 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 patient health status or healthcare delivery that is collected from routine clinical practice rather than from randomized 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 authorize, monitor, or modify the use of medicines in the EU.

Routine Repeated Studies (RR)

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

Study Protocol

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

Very Complex Studies (C4)

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