



Study Protocol

P5-C1-015/P5-C2-012

**DARWIN EU® - Pregnancies amongst
statin users with familial
hypercholesterolaemia**

03/09/2026

Version V2.0

Authors: Guido van Leeuwen, Rana Jajou, Talita Duarte-Salles, Li Luo, Ger Inberg,
Arianna Alamshahi, Ross Williams, Natasha Yefimenko

Public

CONTENTS

LIST OF ABBREVIATIONS.....	5
1. TITLE	7
2. DESCRIPTION OF THE STUDY TEAM	7
3. ABSTRACT	8
4. AMENDMENTS AND UPDATES	10
5. MILESTONES	10
6. RATIONALE AND BACKGROUND	10
7. RESEARCH QUESTION AND OBJECTIVES.....	11
8. RESEARCH METHODS.....	11
8.1. Study design	11
8.2. Follow-up	11
Figure 1. Graphical depiction of the study design for Objective 1, assessing statin use in women of childbearing age (15 – 49) in the study period 2015 – 2025.	12
Figure 2. Graphical depiction of the study design for Objective 2a, assessing statin discontinuation during pregnancy in women of childbearing age (15 – 49), with FH, and a history or at least one pregnancy in the study period 2015 – 2025.	13
Figure 3. Graphical depiction of the study design for Objective 2b, assessing statin discontinuation during breastfeeding in women of childbearing age (15 – 49), with FH, and a history or at least one pregnancy in the study period 2015 – 2025.	14
Figure 4. Graphical depiction of the study design for Objective 2c, assessing statin discontinuation during the first year after birth in women of childbearing age (15 – 49), with FH, and a history or at least one pregnancy in the study period 2015 – 2025.	15
8.3. Study population with inclusion and exclusion criteria.....	15
8.4. Study setting and data sources	17
Table 1. Data sources.....	17
8.5. Study period	18
8.6. Variables	18
8.6.1. Outcome	18
8.6.2. Covariates, including confounders, effect modifiers, and other variables	19
8.7. Study size	20
8.8. Analysis	21
8.8.1. Federated network analyses	21
8.8.2. Data privacy protection	21
8.8.3. Statistical model specification and assumptions of the analytical approach considered	21
Objective 1.....	21
Objective 2.....	21
Table 2. Summary of study objectives and methods.	22
8.8.4. Output	23
MOCK Table 1. General descriptives of cohorts in Objective 1, by data source.	24
MOCK Table 2. Statin use among women of childbearing age (15 – 49) within the study period (2015 – 2025) (Objective 1), by data source.....	25
MOCK Table 3. General descriptives of cohorts in Objective 2, by data source.	25

MOCK Table 4. Statin discontinuation among women of childbearing age (15 – 49), with FH, and history of at least one pregnancy within the study period (2015 – 2025) (Objective 2), by data source.	26
MOCK Figure 1. Statin discontinuation during pregnancy/breastfeeding/during first year after birth as survival analysis among women of childbearing age (15 – 49), with FH, and history of at least one pregnancy within the study period (2015 – 2025) (Objective 2), by data source.	27
8.9. Evidence synthesis.....	28
9. STRENGTHS AND LIMITATIONS	28
10. REFERENCES	29
11. ANNEXES.....	31
ANNEX I. Description of data sources.....	31
Finnish Care Register for Health Care (FinOMOP-THL)	31
The Information System for Research in Primary Care (SIDIAP)	32
Clinical Practice Research Datalink GOLD (CPRD GOLD)	34
IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)	35
IQVIA US - PharMetrics Plus (IQVIA US - PMTX+)	36
ANNEX II. Fitness for use assessment.....	38
ANNEX III. Operational and reporting considerations.....	40
ANNEX IV. Concept lists	42
Table S1. Preliminary list of conditions definitions.	42
Table S2. Preliminary list of medicines definitions.....	42
ANNEX V. ENCePP checklist for study protocols	43
ANNEX VI. Glossary.....	49

Study title	DARWIN EU® - Pregnancies amongst statin users with familial hypercholesterolaemia
Protocol version	V2.0
Date	03/09/2026
EUPAS number	EUPAS1000001100
Active Substance	HMG-CoA Reductase Inhibitors (ATC C10AA)
Medicinal Product	Simvastatin Lovastatin Pravastatin Fluvastatin Atorvastatin Cerivastatin Rosuvastatin Pitavastatin
Research question and objectives	<u>Research Question:</u> What is the rate of pregnancies among women using statins for the indication of familial hypercholesterolaemia (FH)? <u>Study Objectives:</u> - To estimate the number and rate of women of childbearing age using statins: - overall - stratified by familial hypercholesterolaemia (FH) status - stratified by number of pregnancies - Among women with FH and a history of at least one pregnancy, to estimate the number and rate of women who discontinued statin treatment: - during pregnancy - during breastfeeding - during the first year after (each) birth
Countries of study	Finland, Spain, the United Kingdom, the United States
Authors	Guido van Leeuwen g.vanleeuwen@darwin-eu.org Rana Jajou r.jajou@darwin-eu.org Talita Duarte-Salles t.duarte@darwin-eu.org Li Luo l.luo@darwin-eu.org Ger Inberg g.inberg@darwin-eu.org Arianna Alamshahi a.alamshahi@darwin-eu.org Ross Williams r.williams@darwin-eu.org Natasha Yefimenko n.yefimenkonosova@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/terms	Description
APOB	Apolipoprotein B
ATC	Anatomical Therapeutic Chemical
CC	Coordination Centre
CI	Confidence interval
CPRD GOLD	Clinical Practice Research Datalink GOLD
CV	Cardiovascular
CDM	Common Data Model
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DRE	Digital Research Environment
DTZ	Data Transfer Zone
DUS	Drug Utilisation Study
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ETL	Extract, Transform, and Load
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
FH	Familial Hypercholesterolaemia
FinOMOP-THL	Finnish Care Register for Health Care
GDPR	General Data Protection Regulation
GP	General Practitioner
ICD	International Classification of Diseases
IDIAPJGol	Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina
IQR	Interquartile range
IQVIA US - AmbEMR	IQVIA US - Ambulatory EMR
IQVIA US - PMTX+	IQVIA US - PharMetrics Plus
IRB	Institutional Review Board
LDL	low-density lipoprotein
LDLR	LDL receptor
N/A	Not applicable
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
Oxford NDORMS	University of Oxford - Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences
PCSK9	Proprotein convertase subtilisin/Kexin type 9
PET	Perinatal Expansion Table

Acronyms/terms	Description
RWD	Real-World Data
SIDIAP	The Information System for Research in Primary Care
SNOMED	Systematized Nomenclature of Medicine
WHO	World Health Organization

1. TITLE

DARWIN EU® - Pregnancies amongst statin users with familial hypercholesterolaemia

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Guido van Leeuwen	Erasmus MC
Co-Principal Investigator	Rana Jajou Talita Duarte-Salles	Erasmus MC
Epidemiologist	Li Luo	Erasmus MC
Data Scientist	Ger Inberg Arianna Alamshahi Ross Williams	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	Names	Data Partner Organisation*
Finnish Care Register for Health Care (FinOMOP-THL)	Toni Lehtonen Gustav Klingstedt Laura Salonen	FinOMOP
The Information System for Research in Primary Care (SIDIAP)	Elena Roel Herranz Andrea Pistillo Laura Granés González	Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
Clinical Practice Research Datalink GOLD (CPRD GOLD)	Antonella Delmestri Marta Pineda Moncusí Hezekiah Omulo Mandickel Kamtengeni	University of Oxford - Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences (Oxford NDORMS)
IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)	Gargi Jadhav Dina Vojinovic Ellen Gerritsen Akram Sharim Mendez Rangel	IQVIA
IQVIA US - PharMetrics Plus (IQVIA US - PMTX+)	Gargi Jadhav Dina Vojinovic Ellen Gerritsen Akram Sharim Mendez Rangel	IQVIA

*Data Partners do not have an investigator role. The role of Data Partners is to execute the analytical code and review and approve outputs using their expert knowledge about their respective data source.

3. ABSTRACT

Rationale and background

Available data suggest that women with familial hypercholesterolaemia (FH) who become pregnant have worse long-term cardiovascular (CV) outcomes than women with FH who do not become pregnant. This might partly reflect repeated or prolonged interruptions of lipid lowering treatment during pregnancy and breastfeeding, although this remains difficult to assess without large real-world datasets with information on pregnancy history, medicine use, and long-term CV outcomes. While FH is of primary interest for the purpose of this study, the findings may also inform future studies exploring long-term CV outcomes in women with a statin indication, with and without pregnancies.

Research objectives

1. To estimate the number and rate of women of childbearing age using statins:
 - a. overall
 - b. stratified by FH status
 - c. stratified by number of pregnancies
2. Among women with FH and a history of at least one pregnancy, to estimate the number and rate of women who discontinued statin treatment:
 - a. during pregnancy
 - b. during breastfeeding
 - c. during the first year after (each) birth

Methods

Study design

This multinational drug utilisation study will use Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)-mapped healthcare data from Finland (FinOMOP-THL), Spain (SIDIAP), the United Kingdom (UK) (CPRD GOLD), and the United States (US) (IQVIA US - AmbEMR and IQVIA US - PMTX+) from 2015 to 2025.

For Objective 1, women of childbearing age (15 – 49 years) with ≥ 365 days of prior observation will be included to estimate statin initiation overall and stratified by familial hypercholesterolaemia (FH) status and pregnancy history.

For Objective 2, women of childbearing age (15 – 49 years) with FH, statin use, and a history of pregnancy will be included to evaluate statin discontinuation during pregnancy, breastfeeding, and the first postpartum year, using a ≥ 90 -day treatment gap definition.

Implausible or overlapping pregnancy episodes will be excluded. Breastfeeding analyses will be limited to data sources with mother-child linkage.

Data sources

1. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
2. Spain: The Information System for Research in Primary Care (SIDIAP)
3. United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)
4. United States: IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)
5. United States: IQVIA US - PharMetrics Plus (IQVIA US - PMTX+)

Study size

No sample size has been calculated, as this is a drug utilisation study which will not test a specific hypothesis. In addition, already collected available data will be used to estimate statin use in women of childbearing age (15 – 49), and statin discontinuation in women of childbearing age (15 – 49), with FH, and a history of at least one pregnancy. Thus, the sample size is driven by the availability of data for individuals with statin use, FH, and pregnancy. Based on a preliminary feasibility assessment, the expected number of person-counts in the data sources included in this study would range overall from 13,600 (IQVIA US - AmbEMR) to 2,875,900 (IQVIA US - PMTX+) for statin use, from 6,300 (CPRD GOLD) to 136,200 (IQVIA US - PMTX+) for diagnosis of FH, and from 409,207 (SIDAP) to 466,615 (CPRD GOLD) for pregnancy events.

Statistical analysis

Descriptive analyses will be conducted using DARWIN EU® packages; cell counts <5 will be suppressed.

Conclusions

This study will generate Real-World Evidence on statin utilisation and discontinuation rates in women of childbearing age, including with a FH diagnosis, during key reproductive periods across Europe and the United States. The findings will help address important evidence gaps regarding statin management around pregnancy and breastfeeding, and support regulatory and clinical decision-making

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
First draft study protocol	19 August 2026
Final Study Protocol	20 September 2026
Creation of Analytical code	September/October 2026
Execution of Analytical Code on the data	October/November 2026
Draft Study Report	November 2026
Final Study Report	November/December 2026

*Planned dates are dependent on obtaining approvals from the internal review boards, code execution by the data sources, and deliverables review rounds.

6. RATIONALE AND BACKGROUND

Familial hypercholesterolaemia (FH) is an inherited condition in which cholesterol levels are markedly elevated from early life onwards, especially for low-density lipoprotein (LDL) cholesterol.[1] It is usually caused by mutations in genes involved in LDL clearance, most commonly the LDL receptor (LDLR) gene, but also apolipoprotein B (APOB) or proprotein convertase subtilisin/Kexin type 9 (PCSK9).[1-3]

Because LDL cholesterol remains high throughout life, individuals with FH have a significantly increased risk of premature atherosclerotic cardiovascular (CV) disease, such as coronary heart disease, often occurring decades earlier than in the general population. Clinically, FH may be associated with physical signs, such as tendon xanthomas, xanthelasma, and corneal arcus, although many patients are asymptomatic.

FH can be inherited in a heterozygous form or the more severe homozygous form, which leads to significantly elevated cholesterol levels and early CV complications in childhood or adolescence. Early diagnosis and treatment of both heterozygous and homozygous carriers, typically with lifestyle measures and lipid-lowering therapies, such as statins are essential to reduce CV risk.

As a result, affected individuals have a substantially increased lifetime risk of CV disease, including premature coronary artery disease. Statins are therefore a cornerstone of long-term preventive treatment.

Pregnancy is relevant in this context for two reasons. First, pregnancy itself may be associated with differences in long-term CV risk when compared with women who do not become pregnant.[4] Second, for women with FH or another long-term statin indication, pregnancy often leads to temporary discontinuation of statin treatment. This usually concerns both the pregnancy itself and the subsequent period of breastfeeding.[5, 6] For women with one or more pregnancies, these periods of treatment interruption may accumulate over the life course. As such, this may significantly increase the risk of adverse CV outcomes, as treatment interruption and the long-term CV effects of pregnancy may interact rather than simply add up.

While FH is of primary interest for the purpose of this study, this may apply to other long-term indications for statin use in women of childbearing age. Findings from this study may inform future studies exploring long-term CV outcomes in women with a statin indication, with and without pregnancies.

This study makes use of the Perinatal Expansion Table (PET), this was part of the development of the Observational Medical Outcomes Partnership (OMOP) perinatal expansion to address the limited use of OMOP Common Data Model (OMOP CDM) for perinatal research. The expansion includes two main

components: a pregnancy table and an infant table, which contain a set of variables related to each pregnancy episode and newborn infant, respectively. These tables are generated during the mapping process, also known as Extract, Transform, and Load (ETL) process, where source data is mapped into the standardised format required within the OMOP CDM.[7]

7. RESEARCH QUESTION AND OBJECTIVES

Research Question:

What is the rate of pregnancies among women using statins for the indication of familial hypercholesterolaemia (FH)?

Research objectives

3. To estimate the number and rate of women of childbearing age using statins:
 - a. overall
 - b. stratified by FH status
 - c. stratified by number of pregnancies
4. Among women with FH and a history of at least one pregnancy, to estimate the number and rate of women who discontinued statin treatment:
 - a. during pregnancy
 - b. during breastfeeding
 - c. during the first year after (each) birth

8. RESEARCH METHODS

8.1. Study design

A cohort study will be conducted using routinely collected health data from 5 data sources from 3 countries in Europe and the United States (US). This is a drug utilisation study (DUS), which will be conducted to address Objective 1 and Objective 2, assessing statin use and statin discontinuation.

Due to the absence of data mapped to the PET in their respective data source, FinOMOP-THL, IQVIA US - AmbEMR, and IQVIA US - PMTX+ will only contribute to Objective 1a and Objective 1b. Moreover, as CPRD GOLD does not have mother-child linkage[8] and thus does not have information on breastfeeding, it will not contribute to Objective 2b.

8.2. Follow-up

Objective 1

For Objective 1, follow-up will start on the date on which the individual reaches childbearing age (15 – 49 years) at the start or within the study period (01/01/2015 to 31/12/2025) (index date). For Objective 1a and Objective 1b, follow-up will end on the earliest of 1) reaching age >49 within the study period, 2) loss to follow-up, 3) death 4) end of the observation, 5) end of the study period. For Objective 1c, follow-up will end on the earliest of 1) reaching age >49 within the study period, 2) initiation of a new pregnancy episode, 3) loss to follow-up, 4) death, 5) end of the observation, 6) end of the study period (**Figure 1**).

For Objective 1c, individuals will be allowed to contribute to multiple groups (0 pregnancy, 1 pregnancy, or >1 pregnancies) during the study period, provided they meet the criteria specified in **Section 8.3**. For example, if an individual was never pregnant before at their first cohort entry in 2015, they will contribute to the 0 pregnancy group, and follow-up would start. Subsequently, at their first pregnancy in 2017, they

would switch to the 1 pregnancy group, and follow-up would start from that moment on. This makes it possible for individuals to contribute to multiple categories.

Individuals will be followed up until the end of follow-up to assess certain covariates for characterisation, as described in [Section 8.6.2](#).

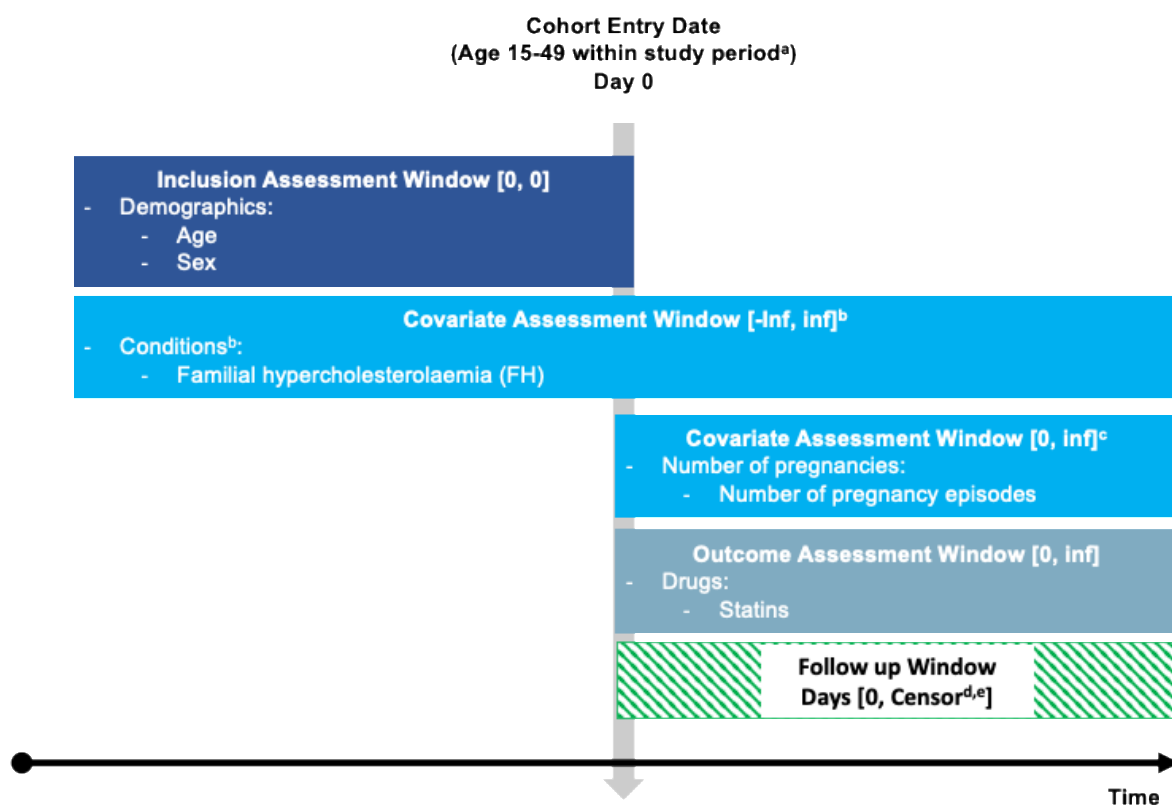


Figure 1. Graphical depiction of the study design for Objective 1, assessing statin use in women of childbearing age (15 – 49) in the study period 2015 – 2025.

- The source population will include women of childbearing age (15 – 49) during the study period (2015 – 2025). Cohort entry will be the moment a woman reaches childbearing age (15 – 49) within the study period.
- Only assessed in Objective 1b.
- Only assessed in Objective 1c. Number of pregnancies is defined as the number of pregnancy episodes recorded in the Perinatal Expansion Table (PET).
- Objectives 1a – b: defined as the earliest of 1) reaching age >49 within the study period, 2) loss to follow-up, 3) death, 4) end of the observation, 5) end of the study period.
- Objective 1c: defined as the earliest of 1) reaching age >49 within the study period, 2) initiation of a new pregnancy episode, 3) loss to follow-up, 4) death, 5) end of the observation, 6) end of the study period.

Objective 2a

For Objective 2a, follow-up will start on the pregnancy start date (index date). Follow-up will end on the earliest of 1) statin discontinuation, 2) pregnancy end date, 3) loss to follow-up, 4) death 5) end of the observation, 6) end of the study period ([Figure 2](#)).

Individuals will be allowed to contribute with multiple entries, corresponding to each pregnancy recorded during the study period, provided they meet the criteria specified in [Section 8.3](#).

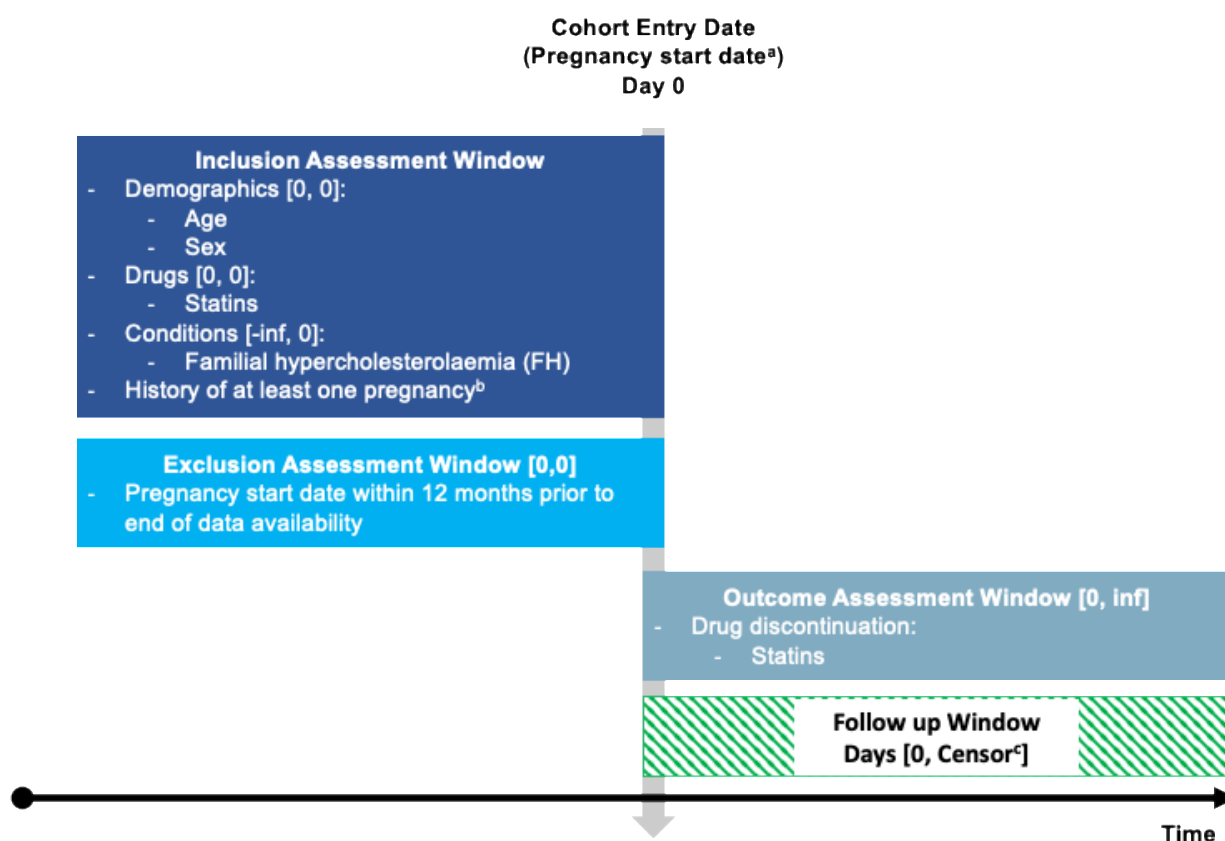


Figure 2. Graphical depiction of the study design for Objective 2a, assessing statin discontinuation during pregnancy in women of childbearing age (15 – 49), with FH, and a history or at least one pregnancy in the study period 2015 – 2025.

- The source population will include women of childbearing age (15 – 49), with FH, statin use, and a history of at least one pregnancy during the study period (2015 – 2025). Cohort entry will be the pregnancy start date, as recorded in the Perinatal Expansion Table (PET).
- As recorded in the PET.
- Defined as the earliest of 1) statin discontinuation, 2) pregnancy end date, 3) loss to follow-up, 4) death 5) end of the observation, 6) end of the study period.

Objective 2b

For Objective 2b, follow-up will start on the pregnancy end date (index date). Follow-up will end on the earliest of 1) statin discontinuation, 2) end of breastfeeding, 3) 12 months of follow-up, 4) loss to follow-up, 5) death 6) end of the observation, 7) end of the study period (**Figure 3**).

Breastfeeding is recorded in the child's Electronic Health Record (EHR) data within the observation table. Subsequently, mother-child linkage, which is available through the fact_relationship table is needed to link the breastfeeding episode in the infants to the mother.[8] The fact_relationship table contains records about the relationship between records stored in any table in the CDM, i.e. a pregnancy/mother (PET) and a child (infant table). The breastfeeding end date is defined as the last breastfeeding registration in the 12 months after the pregnancy end date. In SIDIAP, breastfeeding is generally recorded in routine paediatric assessments during the first 12 months of life.

Individuals will be allowed to contribute with multiple entries, corresponding to each pregnancy recorded during the study period, provided they meet the criteria specified in **Section 8.3**.

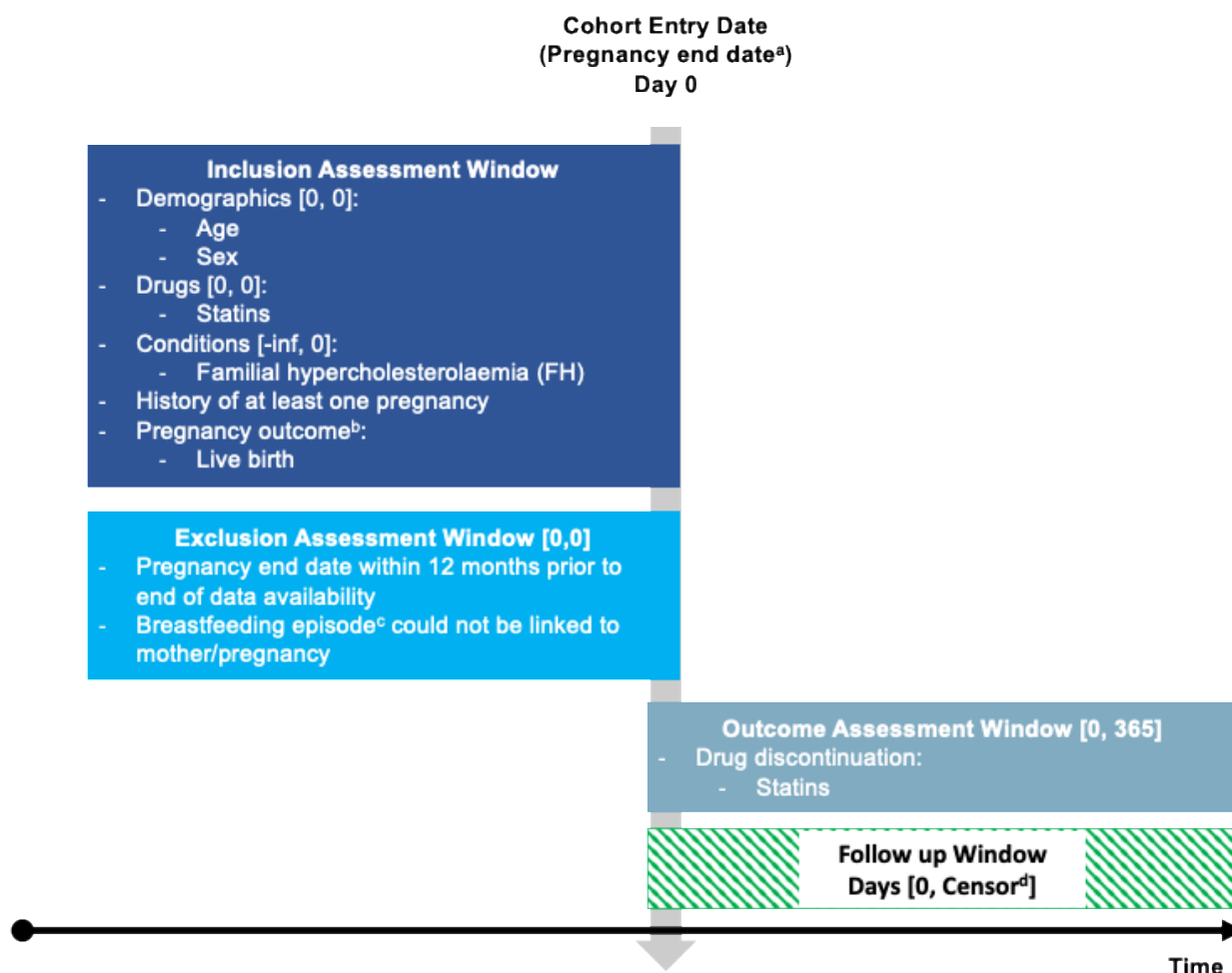


Figure 3. Graphical depiction of the study design for Objective 2b, assessing statin discontinuation during breastfeeding in women of childbearing age (15 – 49), with FH, and a history or at least one pregnancy in the study period 2015 – 2025.

- The source population will include women of childbearing age (15 – 49), with FH, statin use, and a history of at least one pregnancy during the study period (2015 – 2025). Cohort entry will be the pregnancy end date, as recorded in the Perinatal Expansion Table (PET).
- As recorded in the PET.
- As recorded in the Perinatal Expansion Infant Table.
- Defined as the earliest of 1) statin discontinuation, 2) end of breastfeeding, 3) 12 months of follow-up, 4) loss to follow-up, 5) death 6) end of the observation, 7) end of the study period.

Objective 2c

For Objective 2c, follow-up will start on the pregnancy end date (index date). Follow-up will end on the earliest of 1) statin discontinuation, 2) 12 months of follow-up, 3) loss to follow-up, 4) death 5) end of the observation, 6) end of the study period (**Figure 4**).

Individuals will be allowed to contribute with multiple entries, corresponding to each pregnancy recorded during the study period, provided they meet the criteria specified in **Section 8.3**.

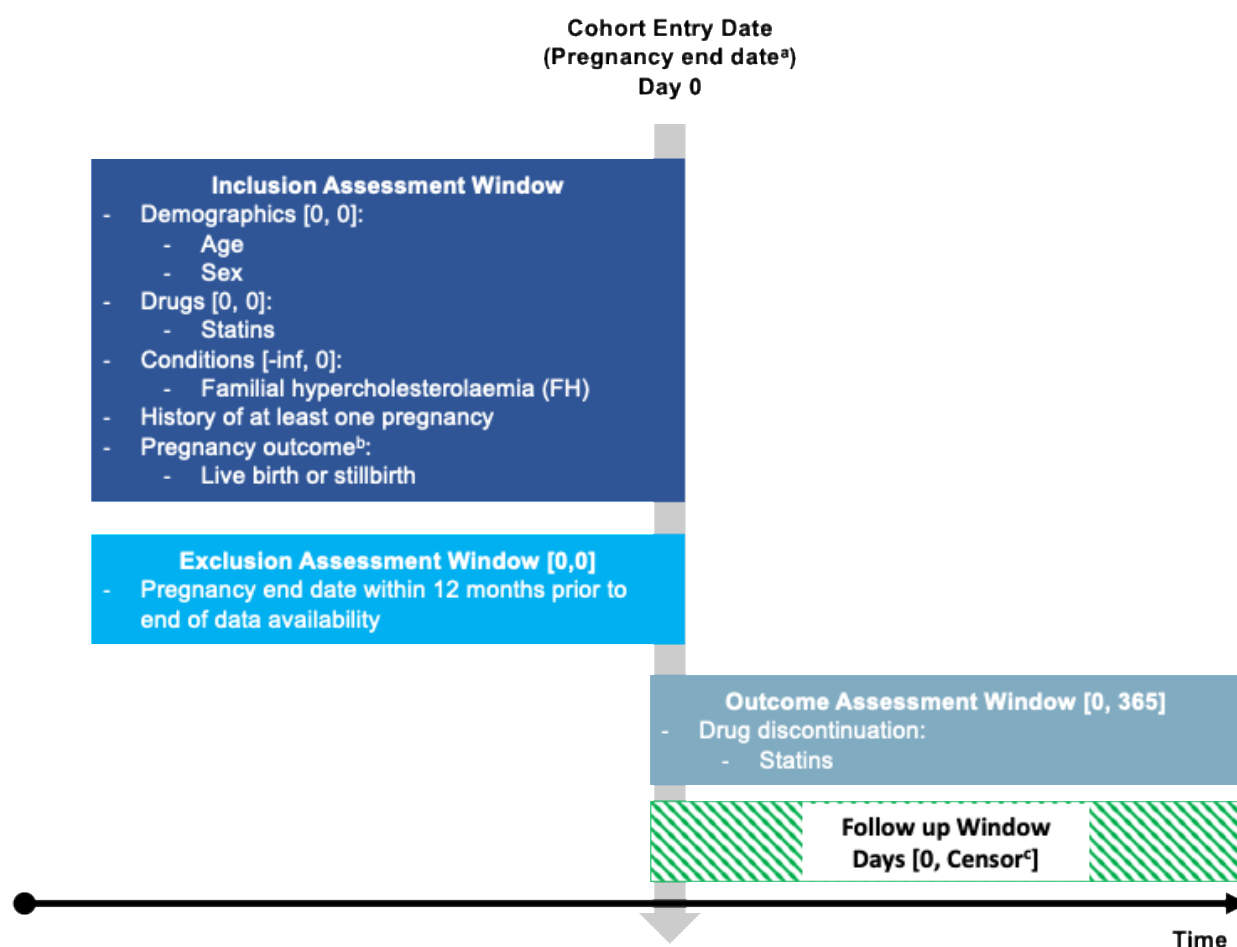


Figure 4. Graphical depiction of the study design for Objective 2c, assessing statin discontinuation during the first year after birth in women of childbearing age (15 – 49), with FH, and a history or at least one pregnancy in the study period 2015 – 2025.

- a. The source population will include women of childbearing age (15 – 49), with FH, statin use, and a history of at least one pregnancy, during the study period (2015 – 2025). Cohort entry will be the pregnancy end date, as recorded in the Perinatal Expansion Table (PET).
- b. As recorded in the PET.
- c. Defined as the earliest of 1) statin discontinuation, 2) 12 months of follow-up, 3) loss to follow-up, 4) death 5) end of the observation, 6) end of study the period.

8.3. Study population with inclusion and exclusion criteria

For Objective 1, the study population will consist of women of childbearing age (15 – 49) within the study period (2015 – 2025).

Objective 1:

Inclusion criteria

- Present in the study period of 01/01/2015 to 31/12/2025 (or the latest available date)
- A minimum of 365 days of available history before the index date
- Women
- Aged 15 to 49 at the index date

For Objective 2, the study population will consist of women of childbearing age (15 – 49), with FH, statin use, and a history of at least one pregnancy within the study period (2015 – 2025).

Objective 2a:

For Objective 2a, all pregnancy episodes recorded for women aged 15 to 49 years, with FH, statin use, and a history of at least one pregnancy will be included. Therefore, individuals may contribute multiple episodes if more than one pregnancy was recorded during the study period.

Inclusion criteria

- Present in the study period of 01/01/2015 to 31/12/2025 (or the latest available date)
- A minimum of 365 days of available history before the index date
- Women
- Aged 15 to 49 at the index date
- A diagnosis of FH
- Using statins at the index date
- A history of at least one pregnancy

Exclusion criteria

- The index date (pregnancy start date) within 12 months prior to end of data availability

Objective 2b:

For Objective 2b, all breastfeeding episodes recorded for infants with mother-child linkage will be described in women aged 15 to 49 years, with FH, statin use, and a history of at least one pregnancy. Therefore, individuals may contribute multiple episodes if breastfeeding was recorded for more than one infant during the study period.

Inclusion criteria

- Present in the study period of 01/01/2015 to 31/12/2025 (or the latest available date)
- A minimum of 365 days of available history before the index date
- Women
- Aged 15 to 49 at the index date
- A diagnosis of FH
- Using statins at the index date
- A history of at least one pregnancy
- Pregnancy outcome of live birth

Exclusion criteria

- The index date (pregnancy end date) within 12 months prior to end of data availability
- Breastfeeding episode could not be linked to a mother/pregnancy

Objective 2c:

For Objective 2c, all pregnancy episodes recorded for women aged 15 to 49 years, with FH, statin use, and a history of at least one pregnancy will be included. Therefore, individuals may contribute multiple episodes if more than one pregnancy was recorded during the study period.

Inclusion criteria

- Present in the study period of 01/01/2015 to 31/12/2025 (or the latest available date)
- A minimum of 365 days of available history before the index date
- Women
- Aged 15 to 49 at the index date
- A diagnosis of FH
- Using statins at the index date
- A history of at least one pregnancy
- Pregnancy outcome of live birth or stillbirth

Exclusion criteria

- The index date (pregnancy end date) within 12 months prior to end of data availability

All objectives including pregnancy or breastfeeding (1c, 2):

Prior quality checks will be performed to ensure the plausibility and observability of the pregnancies recorded. Pregnancies with the following characteristics will be excluded:

- Pregnancies among individuals not observed in the data source at the pregnancy start or end date
- Pregnancies with implausible duration (i.e. pregnancy end date < pregnancy start date, <1 day, or >44 weeks of gestation)
- Pregnancies with conflicting dates (i.e. distinct pregnancies with overlapping dates)

8.4. Study setting and data sources

This study will be conducted using routinely collected data from 5 data sources in the DARWIN EU® Network of Data Partners from 3 countries in Europe and in the US. All data were *a priori* mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). Of these, CPRD GOLD and SIDIAP have mapped data to the PET.[7]

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	Finnish Care Register for Health Care (FinOMOP-THL)	Primary care, secondary care, hospital inpatient care	Registries, EHR	6.15M	2015 – 2026	Objectives 1a, 1b

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to
Spain	The Information System for Research in Primary Care (SIDIAP)	Primary care, hospital inpatient care	EHR	6.07M	2006 – 2024	All Objectives
United Kingdom	Clinical Practice Research Datalink GOLD (CPRD GOLD)	Primary care	EHR	2.37M	1987 – 2025	Objective 1, 2a, and c
United States	IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)	Primary care	EHR	9.13M	2006 – 2026	Objectives 1a, 1b
United States	IQVIA US - PharMetrics Plus (IQVIA US - PMTX+)	Secondary care, hospital inpatient care	Claims	29.51M	2016 – 2025	Objectives 1a, 1b

EHR=Electronic Health Records.

Data sources selection

These data sources fulfil the criteria required in terms of data quality, completeness, timeliness, and representativeness for DUS while covering different regions of the world ([Annex II](#)).

All the included data sources have data on FH and statin use. Of these, SIDIAP and CPRD GOLD have pregnancy-related data mapped to the PET, which will be used in the context of this study. Of the two data sources with data mapped to the PET, only SIDIAP has mother-child linkage and information on breastfeeding. All data sources included have previously participated, or are currently participating, in multi-data source studies on pregnancy using the PET, including the DARWIN EU® PeriNet study.[9-12]

Importantly, in SIDIAP and CPRD GOLD the PET is derived from primary care data, including reproductive health clinics in primary care in SIDIAP. In CPRD GOLD, pregnancy episodes are inferred based on the algorithm developed by Matcho *et al.*[13]

8.5. Study period

The study period is from 01/01/2015 to 31/12/2025 or the most recent data available for each contributing data source. It should be noted that in IQVIA US - PMTX+ data availability begins in 2016; therefore, the study period for this data source was restricted to from 2016 onwards for all objectives.

8.6. Variables

8.6.1. Outcome

Objective 1:

The outcome for this objective will be as follows:

- Use of statins [0, Inf], defined as the first prescription during follow-up
 - Objective 1b: first prescription during follow-up in women with FH
 - Objective 1c: first prescription during follow-up in women with (0/1/>1 pregnancy)

Objective 2:

The outcome for this objective will be as follows:

- Discontinuation of statins [0, inf], defined as the discontinuation of the prescription during follow-up
 - Drug era gap of 90 days, to account for long-term use, where prescriptions generally have a duration of 3 months.

The preliminary concept sets used for the identification of outcomes are described in [Annex IV](#). These codes will be refined during study execution, following the standard DARWIN EU® phenotyping processes, which involve the review of code lists by clinical experts and the review of phenotypes after their execution in the participating data sources.

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

All objectives:

The covariates for all objectives will be as follows:

- Sex
 - Women
- Age
 - 15 to 49 years

Objective 1b:

The covariates for this objective will be as follows:

- Familial hypercholesterolaemia (FH)
 - Assessment window at [-Inf, inf]
 - Stratification:
 - Yes/no

Objective 1c:

The covariates for this objective will be as follows:

- Number of pregnancies (recorded pregnancy episodes in PET), defined at start of follow-up.
 - Assessment window at the index date: [0, Inf]
 - Stratification:
 - 0/1/>1

Objective 2:

The covariates for this objective will be as follows:

- Statin use
 - Assessment window at the index date: [0,0]

- Familial hypercholesterolaemia (FH), defined as a history of FH at start of follow-up.
 - Assessment window at $[-\text{Inf}, 0]$
- History of at least one pregnancy, defined at start of follow-up.
 - Defined as presence in PET

Objective 2a:

The extra covariate for this specific objective will be as follows:

- Pregnancy start date (index date)
 - In SIDIAP: corresponds to the date of the last menstrual period, which is, in most cases, adjusted based on ultrasound results and recorded in the electronic health record in the PET.
 - In CPRD: corresponds to the date of the last menstrual period or with the date of an antenatal visit in primary care.

Objective 2b:

The extra covariates for this specific objective will be as follows:

- Pregnancy outcome, defined at start of follow-up.
 - Assessment window at the index date: $[0,0]$
 - Live birth
- Pregnancy end date (index date)
 - Based on the date when pregnancy ended (based on date of pregnancy outcome)

Objective 2c:

The extra covariates for this specific objective will be as follows:

- Pregnancy outcome, defined at start of follow-up.
 - Assessment window at the index date: $[0,0]$
 - Live birth or still birth
- Pregnancy end date (index date)
 - Based on the date when pregnancy ended (based on date of pregnancy outcome)

The preliminary concept sets used for the identification of covariates are described in [Annex IV](#). These codes will be refined during study execution, following the standard DARWIN EU® phenotyping processes, which involve the review of code lists by clinical experts and the review of phenotypes after their execution in the participating data sources.[14]

8.7. Study size

No sample size has been calculated, as this is a drug utilisation study which will not test a specific hypothesis. In addition, already collected available data will be used to estimate statin use in women of childbearing age (15 – 49), and statin discontinuation in women of childbearing age (15 – 49), with FH, and a history of at least one pregnancy. Thus, the sample size is driven by the availability of data for individuals with statins, FH, and pregnancy. Based on a preliminary feasibility assessment, the expected number of person-counts in the data sources included in this study would range overall from 13,600 (IQVIA US -

AmbEMR) to 2,875,900 (IQVIA US - PMTX+) for statin use, from 6,300 (CPRD GOLD) to 136,200 (IQVIA US - PMTX+) for FH, and from 409,207 (SIDIAP) to 466,615 (CPRD GOLD) for pregnancy.

8.8. Analysis

8.8.1. Federated network analyses

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

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

8.8.2. Data privacy protection

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

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

A summary of the study objectives and associated study design elements is shown in [Table 2](#).

The *CohortCharacteristics*[15] package, developed by DARWIN EU®, will be run for both objectives to describe general characteristics for all the cohorts (age (median (interquartile range (IQR)), age distribution (15 – 19, 20 – 29, 30 – 39, 40 – 49)).

Moreover, the drug utilisation analysis in both objectives will be calculated based on OMOP CDM-mapped data using the R package *DrugUtilisation*,[16] developed by DARWIN EU®.

Objective 1

One estimand (number and rate (%) of statin use) will be calculated in all women of childbearing age (15 – 49 years). The estimand will be calculated over the entire study period, and the results will be stratified by (a history of) FH (yes/no) for Objective 1b, and by number of pregnancies (number of pregnancy episodes in PET, 0/1/>1) for objective 1c (see [Section 8.6.2](#)).

Objective 2

One estimand (number and rate (%) of statin discontinuation) will be calculated in all women of childbearing age (15 – 49 years), with FH, statin use, and a history of at least one pregnancy. The summariseDiscontinuationAsSurvival function of the *DrugUtilisation* package will be used.[17] These analyses use survival analysis as a measure of discontinuation, resulting in a number of subjects, number of events (statin discontinuation), and the survival probability (i.e. the probability to not discontinue over time).

As described in [Section 8.6.2](#), statin discontinuation will be estimated during pregnancy in Objective 2a, during breastfeeding in Objective 2b, and during the first year after birth in Objective 2c. To describe the discontinuation of statins after the start of follow-up, a drug era gap of 90 days will be allowed. In the *DrugUtilisation* package, the censorDate was the cohort end date (end of follow-up), as specified in [Section 8.6.2](#).

All analysis

Cell counts <5 will be suppressed when reporting results to comply with the data source's privacy protection regulations.

Regarding missing data, analyses for all objectives will be conducted using available data without imputation. As such, no specific methods will be applied to address missingness, and the absence of a recorded entry (e.g. condition) will be interpreted as the individual not having the event of interest (e.g. outcome).

Table 2. Summary of study objectives and methods.

Objective	Design	Study population	Treatment of interest	Follow-up	Analysis	Stratifications
1. Statin use	DUS	Women of childbearing age (15 – 49 years) within the study period (2015 – 2025)	Statins	- Start: age 15 – 49 within the study period - Objective 1a and Objective 1b end on the earliest of 1) reaching age >49 within the study period, 2) loss to follow-up, 3) death 4) end of the observation, 5) end of the study period. - For Objective 1c, end on the earliest of 1) reaching age >49 within study period, 2) initiation of a new pregnancy episode, 3) loss to follow-up, 4) death, 5) end of the observation, 6) end of the study period.	Statin use during follow-up	- FH (yes/no) - Number of pregnancies (gravity, 0/1/>1)
2a. Statin discontinuation	DUS	Women of childbearing age (15 – 49 years) with FH and a history of at least one pregnancy within the study period (2015 – 2025)	Statins	- Start: pregnancy start date - End on the earliest of 1) statin discontinuation, 2) pregnancy end date, 3) loss to follow-up, 4) death 5) end of the observation,	Statin discontinuation during follow-up	N/A

Objective	Design	Study population	Treatment of interest	Follow-up	Analysis	Stratifications
				6) end of the study period.		
2b. Statin discontinuation	DUS	Women of childbearing age (15 – 49 years) with FH and a history of at least one pregnancy within the study period (2015 – 2025)	Statins	- Start: pregnancy end date - End on the earliest of 1) statin discontinuation, 2) end of breastfeeding, 3) 12 months of follow-up (4) loss to follow-up, 5) death 6) end of the observation, 7) end of the study period.	Statin discontinuation during follow-up	N/A
2c. Statin discontinuation	DUS	Women of childbearing age (15 – 49 years) with FH and a history of at least one pregnancy within the study period (2015 – 2025)	Statins	- Start: pregnancy end date - End on the earliest of 1) statin discontinuation, 2) 12 months of follow-up, 3) loss to follow-up, 4) death 5) end of the observation, 6) end of the study period.	Statin discontinuation during follow-up	N/A

N/A=Not Applicable; DUS=Drug Utilisation Study; FH= Familial Hypercholesterolaemia.

8.8.4. Output

Output will consist of a PDF report, including an executive summary and the following tables and figures:

- Table 1. General descriptives of cohorts in Objective 1, by data source.
- Table 2. Statin use among women of childbearing age (15 – 49) within the study period (2015 – 2025) (Objective 1).
- Table 3. General descriptives of cohorts in Objective 2, by data source.
- Table 4. Statin discontinuation among women of childbearing age (15 – 49), with FH, and history of at least one pregnancy within the study period (2015 – 2025) (Objective 2), by data source.
- Figure 1/2/3. Statin discontinuation during pregnancy/breastfeeding/during first year after birth as survival analysis among women of childbearing age (15 – 49), with FH, and history of at least one pregnancy within the study period (2015 – 2025) (Objective 2), by data source.

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

MOCK Table 1. General descriptives of cohorts in Objective 1, by data source.

			Data source 1	Data source 2	Data source 3	Data source 4	Data source 5
Overall	Overall	Number of records					
		Number of individuals					
		Age (Median (IQR))					
		Age group (15 – 19, 20 – 29, 30 – 39, 40 – 49)					
		FH (yes)	N (%)				
FH	Yes	Number of records					
		Number of individuals					
		Age (Median (IQR))					
		Age group (15 – 19, 20 – 29, 30 – 39, 40 – 49)					
FH	No	Number of records					
		Number of individuals					
		Age (Median (IQR))					
		Age group (15 – 19, 20 – 29, 30 – 39, 40 – 49)					
Number of pregnancies	0	Number of records					
		Number of individuals					
		Age (Median (IQR))					
		Age group (15 – 19, 20 – 29, 30 – 39, 40 – 49)					
Number of pregnancies	1	Number of records					
		Number of individuals					

			Data source 1	Data source 2	Data source 3	Data source 4	Data source 5
		Age (Median (IQR))					
		Age group (15 – 19, 20 – 29, 30 – 39, 40 – 49)					
Number of pregnancies	>1	Number of records					
		Number of individuals					
		Age (Median (IQR))					
		Age group (15 – 19, 20 – 29, 30 – 39, 40 – 49)					

FH=Familial Hypercholesterolaemia; IQR=interquartile range.

MOCK Table 2. Statin use among women of childbearing age (15 – 49) within the study period (2015 – 2025) (Objective 1), by data source.

					Data source 1		Data source 2		Data source 3		Data source 4		Data source 5	
					Number of subjects	Number of statin users	Number of subjects	Number of statin users	Number of subjects	Number of statin users	Number of subjects	Number of statin users	Number of subjects	Number of statin users
Overall	Overall	Statin use	N	N (%)										
FH	Yes													
	No													
Number of pregnancies	0													
	1													
	>1													

FH=Familial Hypercholesterolaemia.

MOCK Table 3. General descriptives of cohorts in Objective 2, by data source.

			Data source 1	Data source 2	Data source 3	Data source 4	Data source 5
Overall	Overall	Number of records					
		Number of individuals					
		Age (Median (IQR))					

			Data source 1	Data source 2	Data source 3	Data source 4	Data source 5
		Age group (15 – 19, 20 – 29, 30 – 39, 40 – 49)					

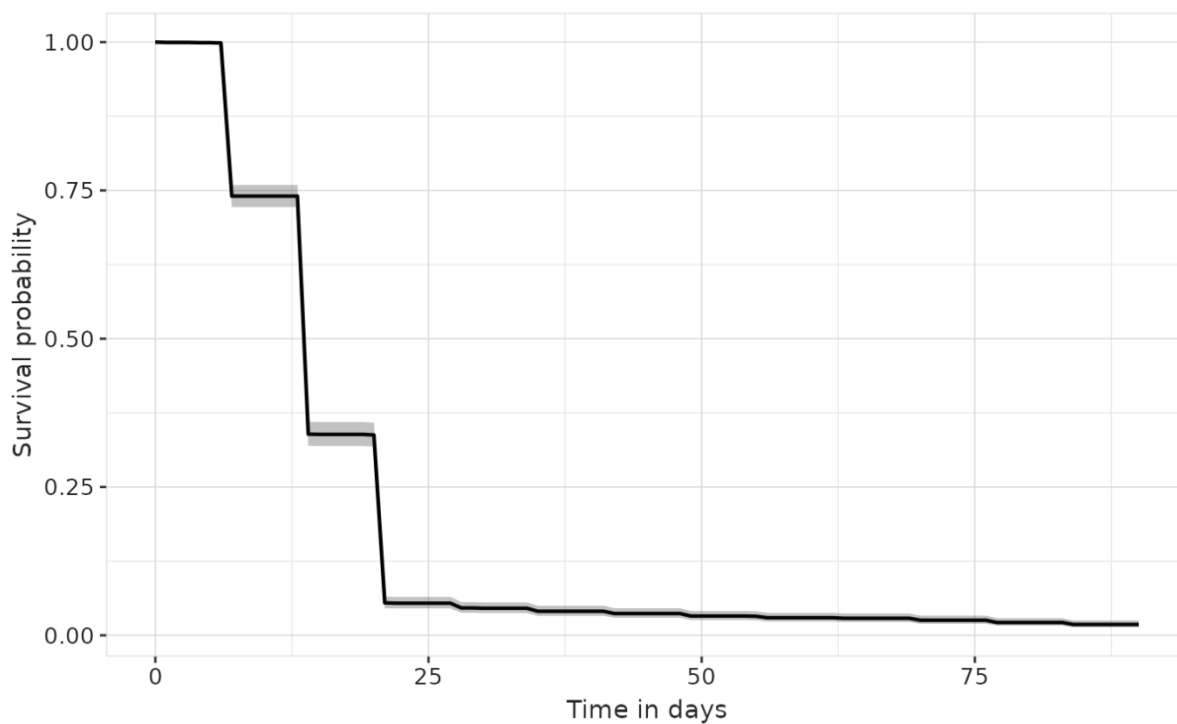
IQR=interquartile range.

MOCK Table 4. Statin discontinuation among women of childbearing age (15 – 49), with FH, and history of at least one pregnancy within the study period (2015 – 2025) (Objective 2), by data source.

		Data source 1	Data source 2
During pregnancy (Obj. 2a)	Number of records	N	
	Number of events (statin discontinuation)	n (%)	
	Median survival (95% CI)*		
	25% quantile (95% CI)		
	75% quantile (95% CI)		
During breastfeeding (Obj. 2b)	Number of records	N	
	Number of events (statin discontinuation)	n (%)	
	Median survival (95% CI)*		
	25% quantile (95% CI)		
	75% quantile (95% CI)		
During first year after birth (Obj. 2c)	Number of records	N	
	Number of events (statin discontinuation)	n (%)	
	Median survival (95% CI)*		
	25% quantile (95% CI)		
	75% quantile (95% CI)		

CI=Confidence interval.

*=In this analysis survival probability can be interpreted as the probability to not discontinue over time.



MOCK Figure 1. Statin discontinuation during pregnancy/breastfeeding/during first year after birth as survival analysis among women of childbearing age (15 – 49), with FH, and history of at least one pregnancy within the study period (2015 – 2025) (Objective 2), by data source.

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

This study has several strengths, namely that the data included is captured in several European countries and the US, covering a large population. This study will utilise linkage with the PET and Pregnancy Expansion Infant Table, which provides the opportunity to gain detailed information and evidence on pregnancies and breastfeeding in women. Moreover, this study will also use standard analytical tools, developed within DARWIN EU®, which provide the opportunity to produce reliable and comprehensive results on a large scale.

The study will be informed by routinely collected health care data and therefore, data quality issues must be considered. In particular, the capture of events of interest may vary across data sources and will be limited to events recorded in the healthcare systems covered by each data source.

Perinatal Expansion Table (PET) specific limitations:

Pregnancies captured will originate from those recorded in the PET, and therefore, the study population will be limited to pregnancies recorded in this table of the OMOP CDM. Based on Objective 4 of the DARWIN EU® PeriNet study, pregnancy outcomes occurring in early gestation are underreported and may not be well captured in the included data sources. The main analysis will include all pregnancies recorded in the data to take early non-live pregnancy outcomes into consideration, as a substantial proportion (15 – 25%) of clinically recognised pregnancies are estimated to end in miscarriage.[10]

The pregnancy start date will not be inferred as part of this study and will be used as recorded in the PET. Its ascertainment will therefore depend on how pregnancy start is recorded or derived in each data source (see [Section 8.4](#)). According to the PET specifications, the pregnancy start date is based on ultrasound estimates or, if ultrasound information is unavailable, calculated from the last menstrual period.[8] Given the nature of routinely collected data, some uncertainty around the exact pregnancy start date cannot be ruled out, and further evaluation of the pregnancy start date recording across participating data sources is outside the scope of this study. To account for these uncertainties, various pregnancy checks are performed, as mentioned in [Section 8.3](#).

Data source specific limitations:

FinOMOP-THL, IQVIA US - AmbEMR, and IQVIA US - PMTX+ do not have available data on detailed pregnancy information for Objectives 1c and 2.

SIDIAP does not include data from private healthcare providers. Therefore, underreporting of health events may occur among individuals who choose to receive antenatal care through private healthcare settings.

For CPRD GOLD, events occurring in hospital settings are likely underreported, as it only covers primary care data. Furthermore, the algorithm used to construct pregnancy episodes in CPRD GOLD does not adequately distinguish between induced and spontaneous abortions. Moreover, the coverage and population size of CPRD GOLD has substantially changed over the last decade, which should be considered when interpreting results stratified by study period. Lastly, CPRD GOLD does not have linkage with the Perinatal Expansion Infant Table, meaning it does not have information on breastfeeding and can therefore not contribute to objective 2b.

10. REFERENCES

1. Bouhairie VE, Goldberg AC. Familial hypercholesterolemia. *Cardiol Clin*. 2015;33(2):169-79.
2. Abifadel M, Elbitar S, El Khoury P, Ghaleb Y, Chémaly M, Moussalli ML, et al. Living the PCSK9 adventure: from the identification of a new gene in familial hypercholesterolemia towards a potential new class of anticholesterol drugs. *Curr Atheroscler Rep*. 2014;16(9):439.
3. Hopkins PN, Toth PP, Ballantyne CM, Rader DJ. Familial hypercholesterolemias: prevalence, genetics, diagnosis and screening recommendations from the National Lipid Association Expert Panel on Familial Hypercholesterolemia. *J Clin Lipidol*. 2011;5(3 Suppl):S9-17.
4. McIlvaine S, Feinberg L, Spiel M. Cardiovascular Disease in Pregnancy. *Neoreviews*. 2021;22(11):e747-e59.
5. Holven KB, Raal FJ, Watts GF, Christensen JJ, Roeters van Lennep J. Challenges in the care of women with familial hypercholesterolaemia during the reproductive period – current evidence and practical guidance. *Atherosclerosis*. 2026;415:120679.
6. Poornima IG, Pulipati VP, Brinton EA, Wild RA. Update on Statin Use in Pregnancy. *Am J Med*. 2023;136(1):12-4.
7. Abellan A, Burn E, Trinh NTH, Burkard T, Callahan A, Fernández-Bertolín S, et al. Expanding the OMOP Common Data Model to Support Perinatal Research in Network Studies. *Pharmacoepidemiol Drug Saf*. 2025;34(2):e70106.
8. Segundo E, Far M, Rodríguez-Casado CI, Elorza JM, Carrere-Molina J, Mallol-Parera R, et al. A mother-child data linkage approach using data from the information system for the development of research in primary care (SIDIAP) in Catalonia. *J Biomed Inform*. 2024;159:104747.
9. Mercade-Besora N, Raventós B, Trinh NT, Li H, Hayati S, Giuliodori A, et al. Effectiveness of COVID-19 mRNA primary and booster vaccination against infection and hospitalisation during pregnancy: a target trial emulation and meta-analysis of data from 4 European countries. *medRxiv*. 2025:2025.05.16.25327593.
10. Politi J D-ST. DARWIN EU® – Utilisation of commonly used benzodiazepines during pregnancy and the incidence of pregnancy losses. EMA; 2025. Report No.: EUPAS1000000536.
11. Prieto-Alhambra D M-BN. Safety of mRNA COVID-19 vaccines during pregnancy. EMA; 2025. Report No.: EUPAS1000000712.
12. EU® D. DARWIN EU® - Enabling pregnancy and mother-child research in DARWIN EU® Data Network - PeriNet. EMA; 2025. Report No.: EUPAS1000000812.
13. Matcho A, Ryan P, Fife D, Gifkins D, Knoll C, Friedman A. Inferring pregnancy episodes and outcomes within a network of observational databases. *PLOS ONE*. 2018;13(2):e0192033.
14. Dernie F, Corby G, Robinson A, Bezer J, Mercade-Besora N, Griffier R, et al. Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU). *Pharmacoepidemiol Drug Saf*. 2024;33(11):e70042.
15. Du M, Prats-Urbe A, Mercadé-Besora N, Lopez-Guell K, Guo Y, Alcalde-Herraiz M, et al. CohortCharacteristics: an R package for population characterisation in observational studies using the OMOP common data model. *European Journal of Epidemiology*. 2026;41(4):507-17.
16. Burkard T, López-Güell K, Gorbachev A, Bellas L, Jödicke AM, Burn E, et al. Calculating daily dose in the Observational Medical Outcomes Partnership Common Data Model. *Pharmacoepidemiology and Drug Safety*. 2024;33(6):e5809.

17. *DrugUtilisation* DE. Summarise discontinuation as a survival analysis [Available from: <https://darwin-eu.github.io/DrugUtilisation/reference/summariseDiscontinuationAsSurvival.html#ref-examples>].
18. Recalde M, Rodríguez C, Burn E, Far M, García D, Carrere-Molina J, et al. Data Resource Profile: The Information System for Research in Primary Care (SIDIAP). *Int J Epidemiol*. 2022;51(6):e324-e36.

11. ANNEXES

ANNEX I. Description of data sources

Finnish Care Register for Health Care (FinOMOP-THL)

#	Section	Description
1	Data source identification and country	FinOMOP-THL (Finnish Care Register for Health Care) Finland
2	Data partner information section	FinOMOP (FinOMOP) The Department of Data and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. The current CDM population comprises all persons having been alive and residing in Finland since the beginning of 2011.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify). THL maintains health registers that cover both public and private, primary, and specialised inpatient, urgent and outpatient health care encounters in Finland, starting from 2011. The entire public sector and private inpatient encounters have been included since 2011, while private outpatient encounters, including occupational care, are included since 2020. Since 1998, the Care Register for Health Care (TerveysHilmo) has covered both public outpatient and inpatient specialized care and private inpatient care. Since 2009, the Finnish National Vaccination Register has covered all vaccinations from the public sector and from a large part of private vaccination providers, with the data coverage from both sections being very good to complete from 2020 onwards. Since 2011, the Register of Primary Health Care Visits (AvoHilmo) has covered public primary care. Since 2020, the register has also covered private outpatient care and occupational care. In addition, the CDM also contains positive COVID-19 test results from the Finnish National Infectious Diseases Register. The register itself covers all laboratory confirmations for around 70 specific microbes from 1995 onwards, but only COVID-19 has currently been mapped to the CDM. The CDM also includes prescription records from multiple different sources. Both Care Register for Health Care and Register of Primary Health Care Visits contain very basic prescription data recorded during health care encounters that include just the ATC-code and trade name of medication. More comprehensive prescription data from Kanta Prescription Centre, maintained by Social Insurance Institution of Finland (Kela), has been integrated into the CDM since its 2024 release. The Kanta Prescription records are based electronic prescriptions, which were adopted by most public health care providers in 2010 and by most private providers by 2017.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. Data is entered by clinicians upon healthcare contact and processed by THL (Kela in the case of Kanta Prescription Centre).
6	General representativeness	The THL data has national coverage and is therefore well representative of the Finnish population. Using the complete population as a basis for the person table also serves to facilitate calculations on a population level, e.g. incidence rates.
7	Data content / source coding	The following coding systems have been OMOP-mapped, typically to a good level of completeness: ICD10fi Finnish Extension, ICPC-2, ATC, Toimenpideluokitus (procedure classification adapted from the Nordic Classification of Surgical Procedures (NCSP)), Terveystieteiden erikoissalat (Hilmo specific provider speciality), Rokotustapa (AR/YDIN National classification for vaccine administration), Tupakointistatus (AR/YDIN National classification for smoking status). Vaccinations are identified on product level based on batch number, trade name, vaccine title, and ATC-code. This is mapped on brand and type in the OMOP CDM.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the

#	Section	Description
		DARWIN EU® data network. Each patient in THL has a unique identifier.
9	Quality control (data source specific)	The source data collection undergoes a structural and semantic validation before entry into the source database. Additionally, some coded variables undergo quality assessment against the respective code systems post entry into the database. The source registers are also assessed for completeness and coverage, with the aim of improving future collection in the areas where data is lacking.
10	Linkage	THL is already a linkage of multiple Finnish registries (see above).
11	Vital status	The National Population registry data forms the basis for forming the patient population. This ensures an up-to-date location (municipality of residence) of patients, as well as complete death occurrences (although not the cause of death).
12	Limitations	All drug records in the CDM are currently based on prescriptions. Kanta Prescription Centre also includes information on drug dispensings, but these have not currently been converted into OMOP CDM. Depending on the type of the medication, a single prescription can be for up to two years dosage of the drug. This means a patient can have up to two-year breaks between observations while actively using the drug. Observation of a prescription also does not mean that the patient necessarily bought or used the medication. The CDM does not currently have meaningful end dates or days of supply for drug exposure. This information is not available for Care Register for Health Care or Register of Primary Health Care Visits, and for Kanta Prescription Centre it is only available in unstructured, free-form format that has not been converted into OMOP CDM. The source system for prescription drug data is only available to the DP until Dec-2023, therefore drug_exposure records from this source are only available in that time range. Some vital signs, like height and weight measurements, are recorded even if the person has not been actually measured, just that the measurement is deemed relevant to the condition or it's treatment. This may lead to duplicate measurements as the value from previous visit might be taken. Not all private health care records are covered for the CDM's entire follow-up time from 2011 onwards. For Register of Primary Health Care Visits and Finnish National Vaccination Register, records from private health care have been available from 2020 onwards. For Kanta Prescription Centre, the coverage of private health care records has been good from 2017 onwards. The inclusion of private health care mainly presents itself as an increase in the number of observations, meaning that it has to be accounted for when interpreting any time series data from the CDM.
13	Main references	Häkkinen, Pirjo; Mölläri, Kaisa; Saukkonen, Sanna-Mari; Väyrynen, Riikka; Mielikäinen, Lasse; Järvelin, Jutta "Hilmo - Sosiaali- ja terveydenhuollon hoitoilmoitus 2020 : Määrittelyt ja ohjeistus : Voimassa 1.1.2020 alkaen" Terveyden ja hyvinvoinnin laitos (2019):
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111187 ; https://catalogues.ema.europa.eu/data-source/1111191 Website: https://thl.fi/en/statistics-and-data/data-and-services/register-descriptions ; https://www.kanta.fi/en/research-and-knowledge-management

The Information System for Research in Primary Care (SIDIAP)

#	Section	Description
1	Data source identification and country	SIDIAP (The Information System for Research in Primary Care) Catalunya, Spain
2	Data partner information section	Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. SIDIAP is a database of primary care electronic health records of the population of Catalonia,

#	Section	Description
		North-East Spain. It contains pseudo-anonymised records of more than 8 million people , of which 6.1 million are active as of 2022, representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – General Practitioner, and hospital inpatient care. SIDIAP captured data includes routine visits, socio-demographics, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, sick leaves and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Catala de la Salut (Catalan Health Institute) is the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content / source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient has a unique id, also upon changing practice or re-registration.
9	Quality control (data source specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g., there are several variables with information concerning the women's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinally and reliability), plausibility (range checks and unusual values), and consistency are inspected through visualisation tools. In addition, before accessing the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences, and prevalence of the data requested for the project, allowing detection of inconsistencies in the data extraction prior to data delivery. External validation processes of the SIDIAP database mainly include assessing the data recorded in SIDIAP through linkage to external GOLD standard data sources, by analysing free text, or by sending questionnaires to health professionals.
10	Linkage	SIDIAP is linked to a hospital discharge database, pharmacy dispensation, and primary care laboratories. It can also be linked to other registries in Catalonia on a project by project basis.
11	Vital status	Mortality is fully captured in SIDIAP. The cause of death is not available, but can be linked to the Spanish death registry on a project by project basis.
12	Limitations	The SIDIAP data is not representative of individuals not using public primary care, and conditions that are usually followed by specialist care might not be properly captured. In addition, there is limited information on lifestyle variables. Patients are followed until Death or when transferring to another primary health care centre that does not contribute to SIDIAP.
13	Main references	Recalde M, Rodríguez C,Burn E,Far M,García D,Carrere-Molina J,Benítez M,Moleras A,Pistillo A,Bolíbar B,Aragón M,Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748

#	Section	Description
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

Clinical Practice Research Datalink GOLD (CPRD GOLD)

#	Section	Description
1	Data source identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) the United Kingdom
2	Data partner information section	University of Oxford - Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences (Oxford NDORMS) NDORMS
3	Coverage and timespan	Data collection since: 1987 Extent: Nation-wide. CPRD GOLD consists of patients in contributing practices using Vision software. Historically this covered the whole of the UK, but the number of contributing practices in the England is dropping. In January 2025 only 3 practices from England were a part of CPRD GOLD, while historical patient data were from the whole of the UK, and will continue to be so. In the future, no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records. Data are entered by clinicians into the EHR. Data is processed by CPRD that provides data releases for research.
6	General representativeness	In the last 10 years, the CPRD GOLD regional distribution of currently contributing general practitioner (GP) practices has significantly shifted, resulting in many new practices joining from Scotland, Wales, and Northern Ireland, and fewer participating from England. These changes have affected the CPRD GOLD population size, regional coverage, and eligibility for data linkages. CPRD GOLD January 2024 contains >21.3 million historical and current patients (12.9 in England, 3.1 in Wales, 4.7 in Scotland, 0.7 in Northern Ireland). Of these, nearly 3 million are currently registered in a GP practice and represent ~4.3% of the estimated current UK population (0.1% in England, 32.3% in Wales, 28.6% in Scotland, 16.2% in Northern Ireland). Patients currently registered in CPRD GOLD January 2024 are broadly representative of the UK population with respect to age and sex. Reference: https://doi.org/10.1093/ije/dyaf077
7	Data content / source coding	Gemsript, Read, dm+d
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In GOLD, a patient can be registered under different ID numbers upon changing practice or re-registration. Researchers are not able to identify these patients, as the data are anonymised. However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (data source specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (UTS). Each practice is associated to an UTS date set when the data quality standards become

#	Section	Description
		satisfactory, and CPRD recommend using only longitudinal data starting from this UTS date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards, and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice UTS date. CPRD also checks data quality standards at the patient level, and associates each patient with a flag, reporting if its data are acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at the Practice Level.
11	Vital status	The date of death in CPRD GOLD has been validated against the Population registry (ONS) mortality data. Reference: https://doi.org/10.1002/pds.4747
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL, Dedman D, Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://www.cprd.com/data/primary-care-data/cprd-GOLD

IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)

#	Section	Description
1	Data source identification and country	IQVIA US - AmbEMR (IQVIA US - Ambulatory EMR) the United States
2	Data partner information section	IQVIA (IQVIA) -
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. Over 100 million patients, with data sourced from more than 800 ambulatory practices and 100,000+ physicians across the United States.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians). The IQVIA Ambulatory EMR-US dataset primarily covers ambulatory care settings, which include: Outpatient physician practices, Primary care and specialty clinics, Independent and group practices, and non-hospital-based care environments.
5	Data collection process	Outpatient electronic health records. Anonymized patient records collected from patient management software used by GPs and selected specialists to document patients' clinical encounters. These records are sent to a central location then processed and aggregated.
6	General representativeness	The data is representative of the US population, with a variety of socioeconomic status patients that is not limited to specific insurances.
7	Data content / source coding	The following source vocabularies are used: CPT, HCPCX, Dx, NDC.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. IQVIA's patient tokenization process generates a consistent, anonymous patient identifier by applying multi-layered, irreversible hashing to encrypted patient attributes, enabling privacy-

#	Section	Description
		protected linkage of de-identified data across care settings without exposing personally identifiable information. This enables longitudinal tracking of patients.
9	Quality control (data source specific)	Integrity constraints including • 230+ IQVIA Internal QC checks • DQD & Achilles review • IQVIA internal business validation check
10	Linkage	Linkage to other IQVIA datasets is possible, e.g. PharMetricsPlus, but has to be done on a project basis.
11	Vital status	There is no linkage to a death registry.
12	Limitations	The main limitation is that no mortality information is available and there is no possibility to link a death registry. Drug dosages cannot always be calculated as for half the drug exposures the quantity is missing.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111118 Website: https://www.iqvia.com/locations/united-states/library/fact-sheets/iqvia-ambulatory-emr-us

IQVIA US - PharMetrics Plus (IQVIA US - PMTX+)

#	Section	Description
1	Data source identification and country	IQVIA US - PMTX+ (IQVIA US - PharMetrics Plus) the United States
2	Data partner information section	IQVIA (IQVIA) -
3	Coverage and timespan	Data collection since: 2016 Extent: Nation-wide. IQVIA PharMetrics® Plus is a comprehensive, longitudinal health plan claims database representing the commercially insured U.S. population.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data. PharMetricsPlus includes adjudicated medical and pharmacy claims from managed care plans. The dataset provides detailed information on diagnoses, procedures, prescription fills, costs, and patient demographics, with broad geographic and payer representation.
5	Data collection process	Insurance/administrative claims. Patients are tokenized and longitudinally tracked over time.
6	General representativeness	National coverage / general population who are privately insured. Includes Commercial, Managed Medicaid and Medicare Advantage. Traditional Medicare and FFS. Public insurance is not included. Representative for all regions and a variety of socioeconomic status patients are part of this data set.
7	Data content / source coding	The source data uses several US codings systems: NDC, ICD9CM, CPT, HCPCS, ICD10CM, ICD9PCS. Data are HIPAA-compliant and include features such as: • Medical and pharmacy claims: inpatient, outpatient, and retail/mail-order prescriptions. • Benefit details: copayments, deductibles, coinsurance. • Provider information: specialty, identifiers. • Longitudinal coverage: millions of individuals with continuous enrolment for three or more years.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients are tokenized and longitudinally tracked over time.
9	Quality control (data source specific)	Integrity constraints including • 230+ IQVIA Internal QC checks • DQD & Achilles review • IQVIA internal business validation check

#	Section	Description
10	Linkage	Linkage capabilities: integration with EMR, lab results, mortality data, and consumer datasets on a project basis.
11	Vital status	Vital status is not available, but linkage to death registry is possible on a project basis.
12	Limitations	Death records are not available. Month and day of birth are not available.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: Website: https://www.iqvia.com/locations/united-states/library/fact-sheets/iqvia-pharmetrics-plus

ANNEX II. Fitness for use assessment

Data source justification for inclusion and key characteristics

Finnish Care Register for Health Care (FinOMOP-THL)

FinOMOP-THL will be included in this study because it is a primary care, secondary care, and hospital inpatient care data source that provides relevant information on familial hypercholesterolaemia and statin use in the women of childbearing age (Objectives 1a, 1b).

Based on a preliminary feasibility assessment, the expected number of person-counts for FH (Objective 1b) and statin use (Objectives 1a, 1b) in FinOMOP-THL will be 16,600 and 429,600, respectively.

Moreover, data availability and follow-up in FinOMOP-THL is sufficient, as data availability starts in 2015, and the date of the most recent data extraction (as of August 2026) is December 2025, which aligns with the study period. The median follow-up of the first observation period in FinOMOP-THL is 4,020 days (Interquartile range (IQR): 4,020 – 4,020).

There are some study specific limitations present in FinOMOP-THL, namely: absence of data on detailed pregnancy information (Objectives 1c, 2).

Lastly, FinOMOP-THL has blanket Institutional Review Board (IRB) approval, which makes the execution of this study feasible within the current study timelines.

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

SIDIAP will be included in this study because it is a primary care data source that provides relevant information on statin use, detailed pregnancy information, and FH in the women of childbearing age (Objective 1) and among women with FH and a history of at least one pregnancy (Objective 2).

Based on a preliminary feasibility assessment, the expected number of person-counts for FH (Objectives 1b, 2), statin use (Objectives 1, 2), pregnancies (Objectives 1c, 2), and breastfeeding (Objective 2b) in SIDIAP will be 10,100, 463,100, 409,207, and 774,300, respectively.

Moreover, data availability and follow-up in SIDIAP is sufficient, as data availability starts in 2006, and the date of the most recent data extraction (as of August 2026) is December 2024, which aligns with the study period. The median follow-up of the first observation period in SIDIAP is 5,760 days (IQR, 2,170 – 6,940).

There are study-specific limitations present in SIDIAP. First, SIDIAP does not include data from private healthcare providers.[18] Therefore, underreporting of health events may occur among individuals who choose to receive antenatal care through private healthcare settings. While reproductive health clinics in primary care are covered, individuals considered to have high-risk pregnancies are referred to outpatient hospital care. Recording of diagnoses and prescribed medications in outpatient hospital care may be underreported, as hospital data is only available for inpatient diagnoses. Lastly, voluntary pregnancy losses may be underreported.

Institutional Review Board (IRB) in SIDIAP approval is expected to take approximately 2 – 3 months, which makes the execution of this study feasible within the current study timelines.

Clinical Practice Research Datalink GOLD (CPRD GOLD)

CPRD GOLD will be included in this study because it is a primary care data source that provides relevant information on statin use, detailed pregnancy information, and FH in the women of childbearing age (Objective 1) and among women with FH and a history of at least one pregnancy (Objectives 2a, 2c).

Based on a preliminary feasibility assessment, the expected number of person-counts for FH (Objectives 1b, 2a, 2c), statin use (Objectives 1, 2a, 2c), and pregnancies (Objectives 1c, 2a, 2c) in CPRD GOLD will be 6,300, 1,079,700, and 466,615, respectively.

Moreover, data availability and follow-up in CPRD GOLD is sufficient, as data availability starts in 1987, and the date of the most recent data extraction (as of August 2026) is November 2025, which aligns with the study period. The median follow-up of the first observation period in CPRD GOLD is 2,160 days (IQR, 729 – 4,940).

There are study-specific limitations present in CPRD GOLD. First, events occurring in hospital settings are likely underreported, as it only covers primary care data. However, given the outcomes assessed in this study, it is expected that most FH cases are also captured in primary care data. Second, the algorithm used to construct pregnancy episodes in CPRD GOLD does not adequately distinguish between induced and spontaneous abortions, and these will therefore likely be affected by misclassification.[13] Third, the coverage and population size of CPRD GOLD has substantially changed over the last decade, which should be considered when interpreting results stratified by study period.

Lastly, CPRD GOLD has blanket IRB approval, which makes the execution of this study feasible within the current study timelines.

IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)

IQVIA US - AmbEMR will be included in this study because it is a primary care data source that provides relevant information on familial hypercholesterolaemia and statin use in the women of childbearing age (Objectives 1a, 1b).

Based on a preliminary feasibility assessment, the expected number of person-counts for FH (Objective 1b) and statin use (Objectives 1a, 1b) in IQVIA US - AmbEMR will be 95,900 and 13,600, respectively.

Moreover, data availability and follow-up in IQVIA US - AmbEMR is sufficient, as data availability starts in 2006, and the date of the most recent data extraction (as of August 2026) is February 2026, which aligns with the study period. The median follow-up of the first observation period in IQVIA US - AmbEMR is 830 days (IQR: 107 – 2,240).

There are some study specific limitations present in IQVIA US - AmbEMR, namely: absence of data on detailed pregnancy information (Objectives 1c, 2).

Lastly, IQVIA US - AmbEMR has blanket IRB approval, which makes the execution of this study feasible within the current study timelines.

IQVIA US - PharMetrics Plus (IQVIA US - PMTX+)

IQVIA US - PMTX+ will be included in this study because it is a secondary care, hospital inpatient care data source that provides relevant information on familial hypercholesterolaemia and statin use in the women of childbearing age (Objectives 1a, 1b).

Based on a preliminary feasibility assessment, the expected number of person-counts for FH (Objective 1b) and statin use (Objectives 1a, 1b) in IQVIA US - PMTX+ will be 136,200 and 2,875,900, respectively.

Moreover, data availability and follow-up in IQVIA US - PMTX+ is sufficient, as data availability starts in 2016, and the date of the most recent data extraction (as of August 2026) is May 2025, which aligns with the study period. The median follow-up of the first observation period in IQVIA US - PMTX+ is 33 days (IQR: 17 – 53).

There are some study specific limitations present in IQVIA US - PMTX+, namely: absence of data on detailed pregnancy information (Objectives 1c, 2).

Lastly, IQVIA US - PMTX+ has blanket IRB approval, which makes the execution of this study feasible within the current study timelines.

ANNEX III. Operational and reporting considerations

DATA MANAGEMENT

Data management

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

The analytic code for this study will be written in R and will use 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, individuals 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® Digital Research Environment (DRE). These output files do not contain any data that allow identification of subjects included in the study. The DRE implements further security measures to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/2016 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 *PhenotypeR* (<https://ohdsi.github.io/PhenotypeR/>) 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 study code will be based on DARWIN EU® R packages: *DrugUtilisation*[16] to characterise drug use, and *CohortCharacteristics*[15] to characterise the cohort by indication. These packages will include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub.

PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

A PDF report including an executive summary, and the specified tables and/or figures will be submitted to EMA by the DARWIN EU® 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. Concept lists

Table S1. Preliminary list of conditions definitions.

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Familial hypercholesterolaemia	Familial hypercholesterolemia	4134862	None	SNOMED
	Carrier of familial hypercholesterolemia	37159710	None	SNOMED
Pregnancy outcome: Live birth (from PET)	Livebirth	4092289	None	SNOMED
Pregnancy outcome: Stillbirth (from PET)	Stillbirth	443213	None	SNOMED
Breastfeeding (from Perinatal Expansion Infant Table)	Maternal exclusive breastfeeding	37166037	None	SNOMED
	Partial maternal breastfeeding	37166036		

Table S2. Preliminary list of medicines definitions.

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
Statins	Simvastatin	HMG-CoA Reductase Inhibitors	C10AA	1539403	Yes
	Rosuvastatin			1510813	
	Pravastatin			1551860	
	Pitavastatin			40165636	
	Lovastatin			1592085	
	Fluvastatin			1549686	
	Cerivastatin			1592180	
	Atorvastatin			1545958	

ANNEX V. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Pregnancies amongst statin users with familial hypercholesterolaemia

EU PAS Register® number: EUPAS1000001100

Study reference number (if applicable): P5-C1-015/P5-C2-012

Section 1: Milestones	Yes	No	N/A	Section Number
1.1 Does the protocol specify timelines for				5
1.1.1 Start of data collection ¹	X	☐	☐	
1.1.2 End of data collection ²	X	☐	☐	
1.1.3 Progress report(s)	☐	☐	X	
1.1.4 Interim report(s)	☐	☐	X	
1.1.5 Registration in the EU PAS Register®	X	☐	☐	
1.1.6 Final report of study results.	X	☐	☐	

Comments:

Section 2: Research question	Yes	No	N/A	Section Number
2.1 Does the formulation of the research question and objectives clearly explain:	☐	☐	☐	6 and 7
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)	X	☐	☐	
2.1.2 The objective(s) of the study?	X	☐	☐	
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	X	☐	☐	
2.1.4 Which hypothesis(-es) is (are) to be tested?	X	☐	☐	
2.1.5 If applicable, that there is no <i>a priori</i> hypothesis?	☐	☐	X	

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.

<u>Section 3: Study design</u>	Yes	No	N/A	Section Number
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	X	☐	☐	8.1/8.2
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	X	☐	☐	8.1/8.2
23.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	X	☐	☐	8.1/8.2
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))	☐	☐	X	
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)	X	☐	☐	8.1/8.2

Comments:

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

Comments:

<u>Section 5: Exposure definition and measurement</u>	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)	X	☐	☐	8.6
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☐	X	
5.3 Is exposure categorised according to time windows?	☐	X	☐	8.6

<u>Section 5: Exposure definition and measurement</u>	Yes	No	N/A	Section Number
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☐	X	☐	8.6
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	X	
5.6 Is (are) (an) appropriate comparator(s) identified?	☐	☐	X	

Comments:

<u>Section 6: Outcome definition and measurement</u>	Yes	No	N/A	Section Number
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	X	☐	☐	8.6
6.2 Does the protocol describe how the outcomes are defined and measured?	X	☐	☐	8.6
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)	☐	☐	X	
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)	☐	☐	X	

Comments:

<u>Section 7: Bias</u>	Yes	No	N/A	Section Number
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	X	☐	☐	8.6/8.8
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	X	☐	☐	8.6/8.8
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	X	☐	☐	8.6/8.8

Comments:

<u>Section 8: Effect measure modification</u>	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)	X	☐	☐	8.6/8.8

Comments:

<u>Section 9: Data sources</u>	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)	X	☐	☐	8.4
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	X	☐	☐	8.4
9.1.3 Covariates and other characteristics?	X	☐	☐	8.4
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)	X	☐	☐	8.4
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	X	☐	☐	8.4
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	X	☐	☐	8.4
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	X	☐	☐	8.4
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☐	X	☐	
9.3.3 Covariates and other characteristics?	X	☐	☐	8.4
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	X	

Comments:

<u>Section 10: Analysis plan</u>	Yes	No	N/A	Section Number
10.1 Are the statistical methods and the reason for their choice described?	X	☐	☐	8.8
10.2 Is study size and/or statistical precision estimated?	X	☐	☐	8.7
10.3 Are descriptive analyses included?	X	☐	☐	8.8

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

Comments:

<u>Section 11: Data management and quality control</u>	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)	X	☐	☐	Annex III
11.2 Are methods of quality assurance described?	X	☐	☐	Annex III
11.3 Is there a system in place for independent review of study results?	X	☐	☐	Annex III

Comments:

<u>Section 12: Limitations</u>	Yes	No	N/A	Section Number
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	X	☐	☐	
12.1.2 Information bias?	X	☐	☐	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).	X	☐	☐	
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)	X	☐	☐	9

Comments:

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

<u>Section 13: Ethical/data protection issues</u>	Yes	No	N/A	Section Number
13.2 Has any outcome of an ethical review procedure been addressed?	X	☐	☐	Annex II/III
13.3 Have data protection requirements been described?	X	☐	☐	Annex II/III

Comments:

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

Comments:

<u>Section 15: Plans for communication of study results</u>	Yes	No	N/A	Section Number
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	X	☐	☐	8.8.4/Annex III
15.2 Are plans described for disseminating study results externally, including publication?	X	☐	☐	8.8.4/Annex III

Comments:

Name of the main author of the protocol: Guido van Leeuwen

Date: 19/08/2026

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

Complex Studies (C3)

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

Coordination Centre (CC)

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

Data Access

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

Data Quality Framework

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

Data Source

A 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.