



Study Protocol

P5-C1-014

DARWIN EU® - Endometriosis and adenomyosis: Incidence, prevalence, diagnosis delay, and treatments in European countries

15/09/2026

Version 3.0

Authors: Dina Vojinovic, Ellen Gerritsen

Public

CONTENTS

LIST OF ABBREVIATIONS.....	5
1. TITLE	7
2. DESCRIPTION OF THE STUDY TEAM	7
3. ABSTRACT	8
4. AMENDMENTS AND UPDATES	11
5. MILESTONES	11
6. RATIONALE AND BACKGROUND	11
7. RESEARCH QUESTION AND OBJECTIVES.....	12
8. RESEARCH METHODS.....	12
8.1. Study design	12
Figure 1. Graphical depiction of the study design (<i>objective 1</i>).....	13
Figure 2. Graphical depiction of the study design (<i>objectives 2–4</i>).....	14
8.2. Follow-up	14
8.3. Study population with inclusion and exclusion criteria.....	15
8.4. Study setting and data sources	17
Table 1. Data sources.....	18
8.5. Study period	19
8.6. Variables	19
8.6.1. Outcome	19
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
Figure 3. Incidence example.....	22
Figure 4. Period prevalence example.....	23
Objectives 2 and 4	23
Objective 3.....	23
8.8.4. Output	24
MOCK Table 1. Study attrition of individuals included in the study population, presented by data source (<i>objective 1</i>).	24
MOCK Table 2. Study attrition of individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, presented by data source (<i>objectives 2 and 4</i>).	24
MOCK Table 3. Study attrition of individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, presented by data source (<i>objective 3</i>).	24
MOCK Table 4. Age at first recorded diagnosis of endometriosis or adenomyosis during 2010–2025 (<i>objective 2</i>).	25
MOCK Table 5. Prespecified symptoms recorded in individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, assessed across predefined time windows (<i>objective 2</i>).	26

MOCK Table 6. Top 10 medications recorded among individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, assessed in predefined time windows (<i>objective 2</i>).....	26
MOCK Table 7. Top 10 procedures recorded among individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, assessed in predefined time windows, in hospital data sources (<i>objective 2</i>) (if feasible).....	27
MOCK Table 8. Time from first recorded prespecified symptom related to endometriosis or adenomyosis to diagnosis and gynaecology-related consultation (<i>objective 3</i>).....	28
MOCK Table 9. Treatments and surgical procedures recorded among individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, assessed in predefined time windows (<i>objective 4</i>).....	29
MOCK Figure 1. Annual incidence rates of endometriosis or adenomyosis diagnosis during the study period (<i>objective 1</i>).....	31
MOCK Figure 2. Annual prevalence of endometriosis or adenomyosis diagnosis during the study period (<i>objective 1</i>).....	31
8.9. Evidence synthesis.....	31
9. STRENGTHS AND LIMITATIONS	32
10. REFERENCES	34
11. ANNEXES.....	35
ANNEX I. Description of data sources.....	35
Hungary: Semmelweis University Clinical Data (SUCD).....	35
Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)	36
Spain: The Information System for Research in Primary Care (SIDIAP)	37
Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)	38
The Netherlands: Integrated Primary Care Information (IPCI).....	39
The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)	41
ANNEX II. Fitness for use assessment.....	43
Table S1. Fitness-for-use assessment of data sources.	45
ANNEX III. Operational and reporting considerations.....	47
ANNEX IV. Lists with concept definitions	49
Table S2. Preliminary list of conditions or procedure definitions.	49
Table S3. Preliminary list of medicines definitions.....	50
Table S4. Preliminary list of gynaecology-related specialty definitions.	52
ANNEX V. Illustrative examples of follow-up and censoring for incidence and prevalence analyses.....	53
Figure S1. Included observation time for the denominator population for incidence rates.	53
Figure S2. Included observation time for the denominator population for prevalence.	54
ANNEX VI. ENCePP checklist for study protocols	55
ANNEX VII. Glossary.....	61

Study title	DARWIN EU® - Endometriosis and adenomyosis: Incidence, prevalence, diagnosis delay, and treatments in European countries
Protocol version	V3.0
Date	15/09/2026
EUPAS number	EUPAS1000001099
Active Substance	Hormonal treatments: combined hormonal contraceptives, progestogens, gonadotrophin releasing hormone (GnRH) agonists, gonadotrophin releasing hormone (GnRH) antagonists, and aromatase inhibitors
Medicinal Product	Not applicable.
Research question and objectives	<u>Research questions</u> What are the epidemiology, diagnostic journey, clinical presentation, and treatment of endometriosis and adenomyosis in European healthcare data sources? <u>Objectives</u> 1. To estimate incidence and prevalence of endometriosis or adenomyosis recording in European healthcare data sources. 2. To describe demographic and clinical characteristics of patients at the time of their first recorded endometriosis or adenomyosis diagnosis, including: • Age at diagnosis • Symptoms related to endometriosis or adenomyosis • Comedication • Procedures (<i>if feasible</i>) 3. To assess diagnosis delay of endometriosis or adenomyosis, measured as time from the first recorded symptom related to endometriosis or adenomyosis to: • The first recorded diagnosis of endometriosis or adenomyosis • The first recorded consultation with a gynaecology-related specialty (including reproductive medicine, benign gynaecology, obstetrics, and gynaecological oncology) (<i>if feasible</i>) 4. To describe treatments for endometriosis or adenomyosis, including pharmacological treatments and surgical procedures.
Countries of study	Hungary, Spain, Sweden, the Netherlands, the United Kingdom
Authors	Dina Vojinovic, d.vojinovic@darwin-eu.org Ellen Gerritsen, e.gerritsen@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/terms	Description
ATC	Anatomical Therapeutic Chemical
CC	Coordination Centre
CI	Confidence Interval
CDM	Common Data Model
CPRD GOLD	Clinical Practice Research Datalink GOLD
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DRE	Digital Research Environment
DTZ	Data Transfer Zone
EHR	Electronic Health Records
EMA	European Medicines Agency
EMC	Erasmus University Medical Center
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ESHRE	European Society of Human Reproduction and Embryology
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
GDPR	General Data Protection Regulation
GnRH	Gonadotrophin-releasing hormone
GP	General Practitioner
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage
ICD	International Classification of Diseases
IDIAPJGol	Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina
IMASIS	Institut Municipal Assistència Sanitària Information System
IPCI	Integrated Primary Care Information
IRB	Institutional Review Board
IQR	Interquartile range
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
RxNorm	Medical prescription normalised
PSMar	Consorci Mar Parc de Salut Barcelona
SIDIAP	The Information System for Research in Primary Care
SMPA-GU	Swedish Medical Products Agency - Gothenburg University
SNOMED	Systematized Nomenclature of Medicine
SU	Semmelweis University
SUCD	Semmelweis University Clinical Data

Acronyms/terms	Description
WHO	World Health Organization

1. TITLE

DARWIN EU® - Endometriosis and adenomyosis: Incidence, prevalence, diagnosis delay, and treatments in European countries.

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Dina Vojinovic Ellen Gerritsen	IQVIA
Data Scientist	Akram Mendez Gargi Jadhav	IQVIA
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	Names	Data Partner Organisation*
Semmelweis University Clinical Data (SUCD)	Kiss Loretta Zsuzsa Székely Orsolya Bagyura Zsolt István	Semmelweis University (SU)
Institut Municipal Assistència Sanitària Information System (IMASIS)	Juan Manuel Ramírez-Anguila Carla Arxé Angela Leis Miguel-Angel Mayer	Consorti Mar Parc de Salut Barcelona (PSMar)
The Information System for Research in Primary Care (SIDAP)	Laura Granés González Agustina Giuliodori Picco Talita Duarte Salles Elena Roel Herranz	Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)	Huiqi Li Fredrik Nyberg Ballin Marcel Rickard Ljung Talbäck Mats	Swedish Medical Products Agency - Gothenburg University (SMPA-GU)
Integrated Primary Care Information (IPCI)	Katia Verhamme Marcel de Wilde Guido van Leeuwen Mees Mosseveld	Erasmus University Medical Center (EMC)
Clinical Practice Research Datalink GOLD (CPRD GOLD)	Antonella Delmestri Marta Pineda Moncusí Hezekiah Omulo	University of Oxford - Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences (Oxford NDORMS)

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

3. ABSTRACT

Title

DARWIN EU® - Endometriosis and adenomyosis: Incidence, prevalence, diagnosis delay, and treatments in European countries.

Rationale and background

Endometriosis is a chronic inflammatory and oestrogen-dependent gynaecological disorder affecting an estimated 5–10% of females of reproductive age. It is associated with a substantial burden on quality of life, including chronic pelvic pain, infertility, and impaired daily functioning. Despite its impact, endometriosis remains underdiagnosed, with delays of up to 10 years between symptom onset and diagnosis. Evidence describing the epidemiology, diagnostic pathways, and treatment patterns of endometriosis across Europe remains limited. Improved understanding of the disease burden and management in routine clinical practice may help inform regulatory decision-making and support better care for those affected.

Research question and objectives

Research questions

What are the epidemiology, diagnostic journey, clinical presentation, and treatment of endometriosis and adenomyosis in European healthcare data sources?

Objectives

The specific objectives of this study are:

1. To estimate incidence and prevalence of endometriosis or adenomyosis recording in European healthcare data sources.
2. To describe demographic and clinical characteristics of patients at the time of their first recorded endometriosis or adenomyosis diagnosis, including:
 - Age at diagnosis.
 - Symptoms related to endometriosis or adenomyosis.
 - Comedication.
 - Procedures (*if feasible*).
3. To assess diagnosis delay of endometriosis or adenomyosis, measured as time from the first recorded symptom related to endometriosis or adenomyosis to:
 - The first recorded diagnosis of endometriosis or adenomyosis.
 - The first recorded consultation with a gynaecology-related specialty (including reproductive medicine, benign gynaecology, obstetrics, and gynaecological oncology) (*if feasible*).
4. To describe treatments for endometriosis or adenomyosis, including pharmacological treatments and surgical procedures.

Methods

Study design

A descriptive cohort study including:

- *Descriptive disease epidemiology*: A population-level cohort analysis will be conducted to estimate incidence and prevalence of endometriosis or adenomyosis (*objective 1*).

- *Characterisation*: A patient-level cohort analysis will be used to describe age, symptoms, procedures (*if feasible*), and comedication (*objective 2*); assess time from first recorded symptoms related to endometriosis or adenomyosis to the first recorded diagnosis of endometriosis or adenomyosis or first recorded consultation with a gynaecology-related specialty (*objective 3*); and describe treatments received for endometriosis or adenomyosis (*objective 4*).

The index date for *objective 1* will be the date on which individuals meet all inclusion and exclusion criteria and become eligible for study participation. The index date for *objectives 2, 3, and 4* will be the date of the first recorded diagnosis of endometriosis or adenomyosis during the study period among females present in the data source during the study period (or the latest available date if data availability ends before 31/12/2025).

Population

Descriptive disease epidemiology (objective 1): The study population will include all females aged 12–59 years present in the data source during the study period (from 01/01/2010 to 31/12/2025) (*prevalent population*). This study population will be restricted to individuals with at least 1 year of data visibility prior to becoming eligible for study inclusion and no recorded diagnosis of endometriosis or adenomyosis prior to the cohort entry (*incident population*).

Characterisation (objectives 2, 3, and 4): The study population will include all females aged 12–59 years with a first recorded diagnosis of endometriosis or adenomyosis during the study period (from 01/01/2010 to 31/12/2025). For *objectives 2 and 4*, at least 1 year of data visibility prior to becoming eligible for study inclusion will be required. For *objective 3*, individuals from primary care data sources and registries will be required to have at least 10 years of data visibility prior to the diagnosis date, while no minimum prior data visibility requirement will be applied to hospital-based data sources. *Objective 3* analyses will be restricted to those with at least one recorded symptom related to endometriosis or adenomyosis prior to diagnosis. Analyses of time from first symptom to first gynaecology-related consultation will be limited to individuals with a recorded symptom, recorded consultation with a gynaecology-related specialty, and a recorded diagnosis of endometriosis.

Variables

Exposure: No primary exposure is assessed.

Outcome: Recorded diagnosis of endometriosis or adenomyosis.

Relevant covariates and other variables: Age, calendar year, symptoms related to endometriosis or adenomyosis, comedication use, procedures, gynaecology-related consultations, and treatments received for endometriosis or adenomyosis.

Data sources

1. Hungary: Semmelweis University Clinical Data (SUCD)
2. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
3. Spain: The Information System for Research in Primary Care (SIDIAP)
4. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
5. The Netherlands: Integrated Primary Care Information (IPCI)
6. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Study size

No sample size has been calculated, as this is a descriptive study which will not test a specific hypothesis. Based on a preliminary feasibility assessment, the expected number of person-counts for endometriosis in the data sources included in this study range from 2,600 (IMASIS) to 54,100 (HI-SPEED).

Statistical analysis

Incidence rates of endometriosis or adenomyosis per 100,000 person-years, along with period prevalence of endometriosis or adenomyosis (*objective 1*), will be estimated in the population of females aged 12–59 years present in the data source during the study period. The choice of denominator will be based on the characteristics of the data source. Prevalence will be estimated for (1) endometriosis, (2) adenomyosis, and for (3) combined endometriosis and/or adenomyosis phenotype, while incidence will only be estimated for the combined endometriosis and/or adenomyosis outcome. Incidence rates will be reported together with 95% Poisson confidence intervals. Prevalence will be expressed as proportions together with binomial 95% confidence intervals. Incidence rates and prevalence will be presented overall and stratified by calendar year and age groups.

Characteristics of individuals with a first recorded diagnosis of endometriosis or adenomyosis during the study period (*objectives 2, 3, and 4*) will be described as follows: categorical variables will be reported as counts and proportions, while continuous variables will be reported as medians with interquartile ranges (IQR). Age (*objective 2*) will be assessed at the date of first recorded diagnosis of endometriosis or adenomyosis (index date). Concomitant medications, procedures, and prespecified symptoms (*objective 2*), as well as pre-specified treatments for endometriosis or adenomyosis (*objective 4*) will be described within specific predefined time windows relative to the index date. Diagnostic delay (*objective 3*) will be evaluated as the time from the earliest recorded occurrence of any prespecified symptom potentially related to endometriosis or adenomyosis to the first recorded diagnosis of endometriosis or adenomyosis, or, where feasible, the first recorded consultation with a gynaecology-related specialty.

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

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	To be confirmed by EMA
Creation of Analytical code	September 2026
Execution of Analytical Code on the data	October 2026
Draft Study Report	November 2026
Final Study Report	4 December 2026

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

6. RATIONALE AND BACKGROUND

Endometriosis is a chronic inflammatory, fibrotic, and oestrogen-dependent gynaecological disorder characterised by the presence of endometrial-like tissue outside the uterus or within the uterine wall (adenomyosis).[1] It affects an estimated 5–10% of females of reproductive age and is one of the leading causes of chronic pelvic pain and infertility. Symptoms vary widely and may include dysmenorrhoea, pelvic or abdominal pain, dysuria, dyspareunia, and infertility, contributing to a substantial burden on quality of life, daily functioning, and mental well-being.[2, 3]

Endometriosis remains underdiagnosed, with delays of up to 10 years between symptom onset and diagnosis.[2, 4] The optimal management remains uncertain as the disease aetiology is not fully understood and there are currently no reliable biomarkers for diagnosis. Treatment options include analgesics, hormonal therapies, gonadotrophin-releasing hormone (GnRH) analogues, and aromatase inhibitors, aiming to control symptoms and reduce the need for surgical intervention.[2, 3]

Although diagnostic delay is commonly defined as the interval between symptom onset and diagnosis, symptom onset is not directly captured in routinely collected healthcare data. Symptoms may only be recorded when individuals seek care and may be incompletely or inconsistently documented. Consequently, measures based on the recorded symptoms reflect healthcare encounters rather than true symptom onset and may therefore differ from the true interval between symptom onset and diagnosis.

Despite its significant burden, evidence describing the epidemiology, diagnostic pathways, and treatment of endometriosis across Europe remains limited.[5] A better understanding of the incidence and prevalence of endometriosis, the time between recorded symptoms and diagnosis, and treatment in routine clinical practice may help inform regulatory decision-making, improve disease awareness, and ultimately support better care for individuals affected by this condition.

7. RESEARCH QUESTION AND OBJECTIVES

Research questions

What are the epidemiology, diagnostic journey, clinical presentation, and treatment of endometriosis and adenomyosis in European healthcare data sources?

Objectives

The specific objectives of this study are:

1. To estimate incidence and prevalence of endometriosis or adenomyosis recording in European healthcare data sources.
2. To describe demographic and clinical characteristics of patients at the time of their first recorded endometriosis or adenomyosis diagnosis, including:
 - Age at diagnosis.
 - Symptoms related to endometriosis or adenomyosis.
 - Comedication.
 - Procedures (*if feasible*).
3. To assess diagnosis delay of endometriosis or adenomyosis, measured as time from the first recorded symptom related to endometriosis or adenomyosis to:
 - The first recorded diagnosis of endometriosis or adenomyosis.
 - The first recorded consultation with a gynaecology-related specialty (including reproductive medicine, benign gynaecology, obstetrics, and gynaecological oncology) (*if feasible*).
4. To describe treatments for endometriosis or adenomyosis, including pharmacological treatments and surgical procedures.

8. RESEARCH METHODS

8.1. Study design

A cohort study will be conducted using routinely collected health data from 6 data sources from 5 countries across Europe. The study will comprise:

- A *descriptive disease epidemiology* study to address *objective 1*, assessing incidence and prevalence of endometriosis or adenomyosis (**Figure 1**).
- A *characterisation* study to address *objectives 2, 3, and 4*, assessing age, prespecified symptoms, procedures, and comedication (*objective 2*), time from first recorded symptom related to endometriosis or adenomyosis to first recorded diagnosis of endometriosis or adenomyosis or first recorded consultation with a gynaecology-related specialty (*objective 3*), and treatments received for endometriosis or adenomyosis (*objective 4*) (**Figure 2**).

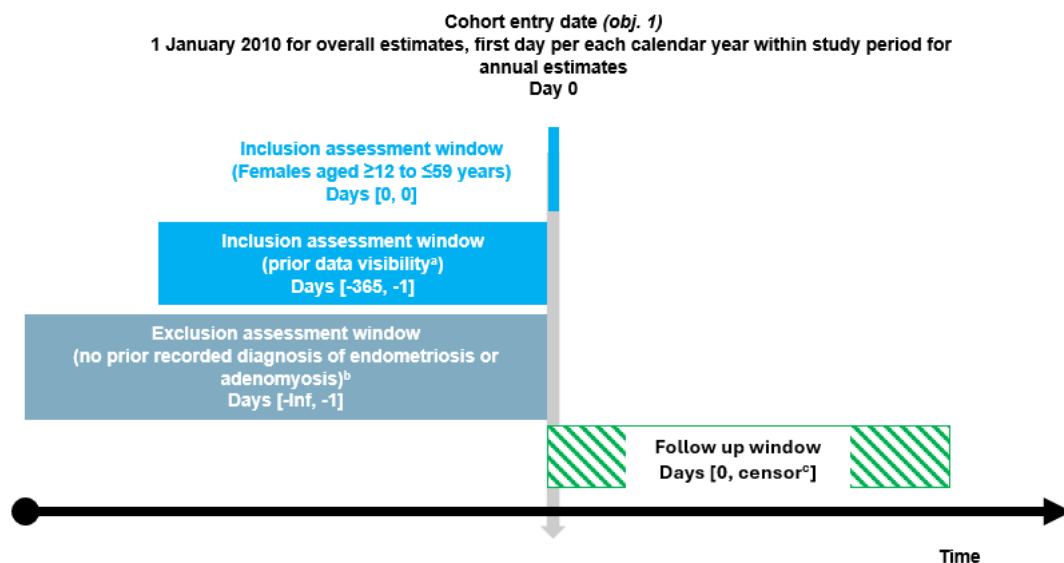


Figure 1. Graphical depiction of the study design (*objective 1*).

- a. Prior data visibility criterion is only applicable for the estimation of incidence rates.
 - b. Exclusion criterion is only applicable for the estimation of incidence rates.
 - c. Follow-up will continue until earliest of: reaching 60 years of age, loss to follow-up, death, end of observation period, or study end date (31/12/2025).
- For the incidence analysis, individuals will additionally be censored at the first recorded diagnosis of endometriosis or adenomyosis, after which they will no longer be considered at risk.

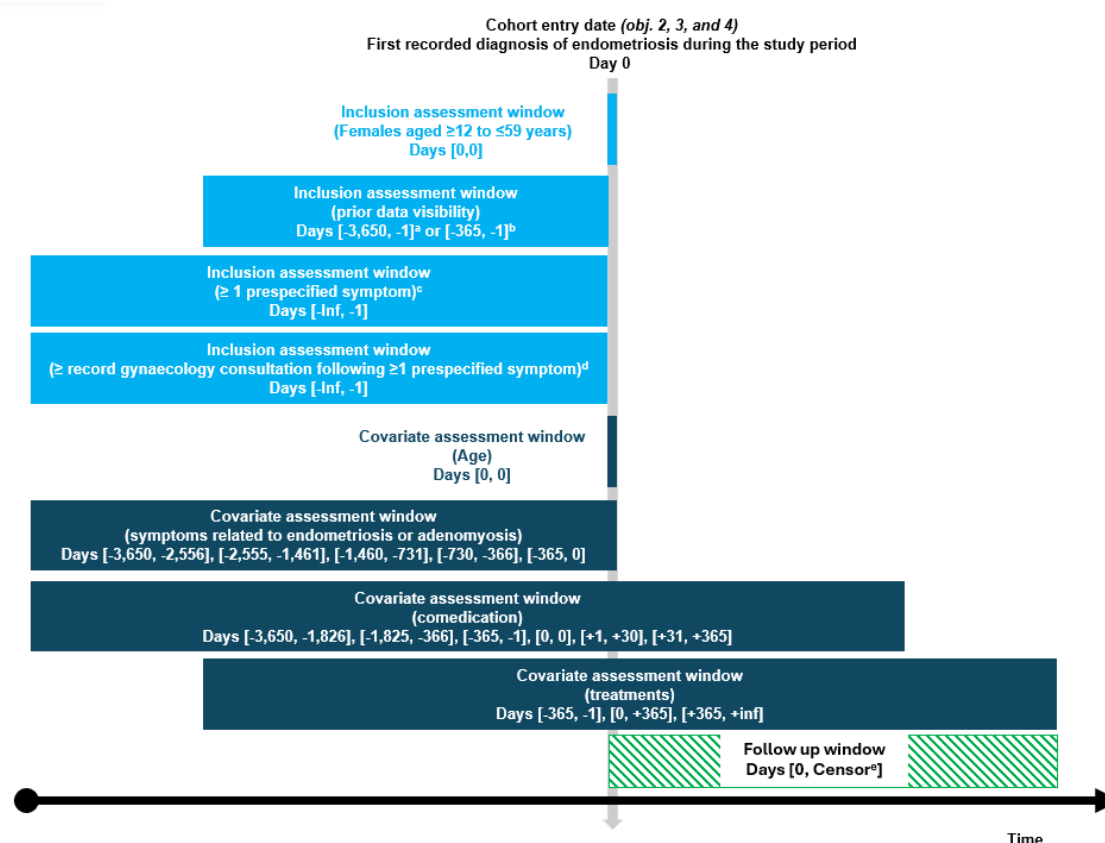


Figure 2. Graphical depiction of the study design (*objectives 2–4*).

- Eligible individuals from primary care data sources and registries are required to have at least 10 years (3,650 days) of data visibility prior to their first recorded diagnosis of endometriosis or adenomyosis (*objective 3*). For hospital-based data sources, no minimum prior data visibility requirement will be applied.
- Eligible individuals are required to have at least 1 year (365 days) of data visibility prior to the first recorded diagnosis of endometriosis or adenomyosis (*objectives 2 and 4*).
- At least one recorded symptom related to endometriosis or adenomyosis before first recorded diagnosis of endometriosis or adenomyosis (*objective 3*). An overview of prespecified symptoms of interest is provided in [Section 8.6.3](#).
- At least one recorded consultation with a gynaecology-related specialty after the first recorded symptom related to endometriosis or adenomyosis and on or before diagnosis (*objective 3*). An overview of prespecified symptoms of interest is provided in [Section 8.6.3](#).
- Follow-up will continue until earliest of: reaching 60 years of age, loss to follow-up, death, end of observation period, or study end date (31/12/2025).

Assessment of symptoms, comedication, procedures, diagnostic delay, and treatments is dependent on the length of available observation time before and after the index date. Consequently, variability in data availability across individuals and data sources may affect the completeness of captured historical events and follow-up information.

8.2. Follow-up

For the calculation of incidence rates of first recorded endometriosis or adenomyosis diagnosis during the study period (*objective 1*), follow-up will start on the respective date of the latest of the following: i) study start date (01/01/2010) or ii) date on which individual has 365 days of prior data visibility available, or iii) date on which individual reaches 12 years of age. Individuals will contribute person-time only while aged ≥ 12 to ≤ 59 years. For prevalence estimation (*objective 1*), follow-up will start on the earliest date within the study period (01/01/2010 – 31/12/2025) on which an individual is recorded in the data source while aged ≥ 12 to ≤ 59 years.

For patient-level characterisation of females with a first recorded diagnosis of endometriosis or adenomyosis (*objectives 2 and 4*), follow-up will start on the date on which individual fulfils all eligibility criteria, including the first recorded diagnosis of endometriosis or adenomyosis occurring between 01/01/2010 and 31/12/2025 (or latest date available, if earlier), while aged ≥ 12 to ≤ 59 years, and having at least 365 days of prior data visibility available.

For the assessment of diagnostic delay (*objective 3*), individuals will become eligible for inclusion on the date they fulfil all study eligibility criteria, including a first recorded diagnosis of endometriosis or adenomyosis occurring during the study period and the required prior data visibility. Individuals from primary care data sources and registries must have at least 3,650 days of data visibility prior to the first recorded diagnosis of endometriosis or adenomyosis, whereas no minimum prior data visibility requirement will be applied for hospital-based data sources. The study population will be restricted to those with at least one recorded symptom related to endometriosis or adenomyosis prior to the first recorded diagnosis of endometriosis or adenomyosis. For analyses of time from first recorded symptom to first consultation with a gynaecology-related specialty, individuals must additionally have at least one recorded consultation with a gynaecology-related specialty after the first recorded symptom. Diagnostic delay will be assessed retrospectively.

Follow-up will end on the earliest of: reaching 60 years of age, loss to follow-up, death, end of the observation period (the latest available data), or the study end date (31/12/2025). For the incidence analysis (*objective 1*), individuals will also be censored at the first recorded diagnosis of endometriosis or adenomyosis, after which they are no longer considered at risk.

The preliminary code lists used to identify diagnoses of endometriosis or adenomyosis and related symptoms, which will define the start and or end of follow-up determining events depending on the objective are presented in [ANNEX IV](#).

Study participants in the denominator population will begin contributing person-time-at-risk from the index date until the earliest censoring event, as described above in [Section 8.2](#). Illustrative examples of denominator population entry, follow-up, person-time contribution, and censoring for incidence and prevalence analyses are provided in [ANNEX V](#).

8.3. Study population with inclusion and exclusion criteria

For the descriptive disease epidemiology study on incidence and prevalence of endometriosis or adenomyosis, the study population will include (*objective 1*):

Incidence population

Inclusion criteria

- Females present in the data source during the study period (01/01/2010 to 31/12/2025 (or the latest available date, if earlier)).
- Age ≥ 12 to ≤ 59 years during the study period.
- At least 365 days of data visibility prior to cohort entry.

Exclusion criteria

- Recorded diagnosis of endometriosis or adenomyosis prior to cohort entry.

Prevalence population

Inclusion criteria

- Females present in the data source during the study period (01/01/2010 to 31/12/2025 (or the latest available date, if earlier)).

- Age ≥ 12 to ≤ 59 years during the study period.

Exclusion criteria

- None.

For the characterisation of demographic and clinical characteristics at first recorded diagnosis of endometriosis or adenomyosis, and the description of treatments among females diagnosed with endometriosis or adenomyosis, the study population will include (*objectives 2 and 4*):

Inclusion criteria:

- First recorded diagnosis of endometriosis or adenomyosis during the study period (01/01/2010 to 31/12/2025 (or the latest available date, if earlier)).
- Females aged ≥ 12 to ≤ 59 years at the date of first recorded diagnosis of endometriosis or adenomyosis.
- At least 365 days of data visibility prior to the first recorded diagnosis of endometriosis or adenomyosis.

Exclusion criteria:

- None.

For the assessment of diagnostic delay, the study population will include (*objective 3*):

Inclusion criteria:

- First recorded diagnosis of endometriosis or adenomyosis during the study period (01/01/2010 to 31/12/2025 (or the latest available date, if earlier)).
- Females aged ≥ 12 to ≤ 59 years at the date of first recorded diagnosis of endometriosis or adenomyosis.
- At least 3,650 days of data visibility prior to the index date for primary care data sources and registries and no minimum prior data visibility requirement for hospital-based data sources.
- At least one recorded prespecified symptom related to endometriosis or adenomyosis before first recorded diagnosis of endometriosis or adenomyosis.
- At least one recorded consultation with a gynaecology-related specialty after the first recorded prespecified symptom, where feasible (for the assessment of time from first recorded symptom to consultation with a gynaecology-related specialty).

Exclusion criteria:

- None.

8.4. Study setting and data sources

This study will be conducted using routinely collected data from 6 primary care, hospital, and registry data sources in the DARWIN EU® Network of Data Partners from 5 European countries ([Table 1, ANNEX I](#)). All data were *a priori* mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

Data sources:

1. Hungary: Semmelweis University Clinical Data (SUCD)
2. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
3. Spain: The Information System for Research in Primary Care (SIDIAP)
4. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
5. The Netherlands: Integrated Primary Care Information (IPCI)
6. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

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	Reliable period for endometriosis/adenomyosis analyses	Contributing to
Hungary	SUCD	Secondary Care	EHR	267,010	2010 – 06/2026	From 2011	All objectives
Spain	IMASIS	Secondary Care	EHR	169,560	1990 – 01/2026	From 2006	All objectives
Spain	SIDIAP	Outpatient General Practitioner Care	EHR	6,068,000	2006 – 12/2024	Entire study period	All objectives
Sweden	HI-SPEED	Registry, Outpatient General Practitioner Care, Inpatient (Hospital) Care	Registry	10,595,600	2015 – 12/2025	Endometriosis/adenomyosis from 2009; drugs from 2018	All objectives
The Netherlands	IPCI	Outpatient General Practitioner Care	EHR	1,339,300	2006 – 12/2025	Entire study period	All objectives
The United Kingdom	CPRD GOLD	Outpatient General Practitioner Care	EHR	2,371,680	1987 – 11/2025	From 2009	All objectives

CPRD GOLD = Clinical Practice Research Datalink; EHR = Electronic Health Record; HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS = Institut Municipal Assistència Sanitària Information System; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care; SUCD = Semmelweis University Clinical Data.

*SUCD and IMASIS are secondary-care data sources that include both specialist ambulatory and hospital outpatient care, as well as inpatient care. This is particularly relevant for endometriosis and adenomyosis, as diagnosis and management may occur in outpatient settings without the need for hospitalisation.

**Number of active individuals is defined as the number of individuals present in the last six months of the respective data source.

Data sources selection

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

8.5. Study period

The study period is from 01/01/2010 to 31/12/2025 or the latest available date if data availability ends before 31/12/2025.

Not all participating data sources will have data available for the entire predefined study period (01/01/2010 to 31/12/2025). Detailed information about the study period covered per data source can be found in [Section 8.4](#) and [Table 1](#).

8.6. Variables

8.6.1. Outcome

The outcome of interest will be first recorded diagnosis of endometriosis or adenomyosis identified using predefined phenotype definitions.

The preliminary concept sets used for the identification of endometriosis or adenomyosis 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 and the review of phenotypes after their execution in the participating data sources. Outcomes will be defined following input from the European Medicines Agency (EMA).

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

The variables for stratification of incidence rates and prevalence (*objective 1*) will be as follows:

- Calendar year
- Age groups (overall, ≥ 12 to ≤ 17 , ≥ 18 to ≤ 29 , ≥ 30 to ≤ 39 , ≥ 40 to ≤ 49 , ≥ 50 to ≤ 59 years)

The variables for characterisation of individuals with first recorded diagnosis of endometriosis or adenomyosis (*objective 2*) will be as follows:

- Age, defined at the index date:
 - Age (median, interquartile range (IQR))
 - Age groups (overall, ≥ 12 to ≤ 17 , ≥ 18 to ≤ 29 , ≥ 30 to ≤ 39 , ≥ 40 to ≤ 49 , ≥ 50 to ≤ 59 years)
- Symptoms related to endometriosis or adenomyosis will be described based on prespecified conditions recorded before the index date, assessment window 1 year prior or on the index date [-365, 0], one to 2 years prior to the index date [-730, -366], 2 to 4 years prior to the index date [-1,460, -731], 4 to seven 7 prior to the index date [-2,555, -1,461], 7 to 10 years prior to the index date [-3,650, -2,556]:
 - Pelvic or lower abdominal pain, including dysmenorrhea
 - Dyspareunia
 - Urinary tract pain
 - Gastrointestinal symptoms, including diarrhoea, constipation, bloating, painful defecation, and bowel dysfunction
 - Excessive menstrual bleeding, including menorrhagia and mid cycle bleeding

- Infertility

Symptom categories will be assessed individually within each assessment window. Categories may be refined following diagnostics results.

- Large scale characterisation: The top 10 most frequently prescribed or dispensed medications (proxy for comedication) and the top 10 most frequently recorded procedures will be identified through large-scale characterisation. These will be assessed 10 years to 5 years before the index date [-3,650, -1,826], 5 years to 1 year before the index date [-1,825, -366], 1 year before the index date [-365, -1], on the index date [0, 0], 1 to 30 days after the index date [+1, +30], 31 to 365 days after the index date [+31, +365]. The top 10 most frequent recorded procedures will only be assessed in hospital data sources.

The variables for the assessment of delay in diagnosis (*objective 3*) will be as follows:

- Any prespecified symptom related to endometriosis or adenomyosis, including pelvic or lower abdominal pain, dyspareunia, urinary tract pain, gastrointestinal symptoms, excessive menstrual bleeding, and infertility

The variables considered for characterisation of treatments recorded among females with endometriosis or adenomyosis (*objective 4*) will be as follows:

- Combined hormonal contraceptives
- Progestogens
- Gonadotrophin releasing hormone (GnRH) agonists
- Gonadotrophin releasing hormone (GnRH) antagonists
- Aromatase inhibitors

Treatment classes have been informed by the European Society of Human Reproduction and Embryology (ESHRE) guideline on endometriosis. Corresponding concept sets will be defined using relevant ingredients identified for each treatment class, informed by the WHO Anatomical Therapeutic Chemical (ATC) classification system, and refined during study execution. Results will be presented at the treatment-level class and ingredient/substance level.

- Surgical procedures relevant to management of endometriosis or adenomyosis
 - Assessment windows for characterisation of treatments will include 1 year before the index date [-365, -1], index date through 1 year after the index date [0, +365], and 1 year to any time after the index date [+366, +inf].

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 and the review of phenotypes after their execution in the participating data sources. Phenotypes will be defined following input from the EMA.

8.7. Study size

No sample size has been calculated, as this is a descriptive disease epidemiology study which will not test a specific hypothesis. In addition, already collected available data will be used to estimate annual incidence and annual prevalence of endometriosis or adenomyosis and to characterise the demographic and clinical characteristics, diagnostic delay, and treatments of affected individuals. Thus, the sample size is driven by the availability of data for individuals. Based on a preliminary feasibility assessment, the expected number of person-counts for endometriosis in the data sources included in this study range from 2,600 (IMASIS) to 54,100 (HI-SPEED).

8.8. Analysis

8.8.1. Federated network analyses

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

Before sharing the study package, test runs of the analytics will be performed on a subset of the data sources, and quality control checks will be performed. After all the tests are passed (see [ANNEX III](#)), 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 a secure online repository of the Data Transfer Zone (DTZ). All results will be locked and timestamped for reproducibility and transparency. The study results of all data sources will be checked, after which they are made available to the team, and the Study Dissemination Phase can start. All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing patient-level data. Cell counts <5 will be suppressed when reporting results to comply with the data source's privacy protection regulations.

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

Objective 1

The incidence and prevalence of endometriosis and adenomyosis will be estimated based on OMOP CDM mapped data using the *IncidencePrevalence* R package (<https://darwin-eu.github.io/IncidencePrevalence/>), developed by DARWIN EU®.

Incidence rates will be estimated only for the combined endometriosis and/or adenomyosis phenotype, whereas prevalence estimates will be estimated separately for (1) endometriosis, (2) adenomyosis, and (3) combined endometriosis and/or adenomyosis phenotype. For the combined incidence outcome, the event date will be defined as the earliest recorded diagnosis of either endometriosis or adenomyosis. Individuals with a recorded diagnosis of either condition before cohort entry will be excluded from the combined incidence population. For the separate prevalence estimates, individuals with recorded diagnoses of both conditions may contribute to both the endometriosis and adenomyosis phenotypes. For the combined prevalence estimate, each individual will be counted only once within each assessment period, irrespective of whether one or both conditions are recorded.

Incidence rates of combined endometriosis or adenomyosis phenotype among females aged ≥ 12 to ≤ 59 years will be calculated as the number of individuals with a first recorded diagnosis of endometriosis or adenomyosis during the study period per 100,000 person-years, among individuals fulfilling the inclusion and exclusion criteria. Incidence rates will be reported together with 95% Poisson confidence intervals. Individuals who enter the denominator population will contribute time-at-risk from cohort entry until the earliest of the following: aged 60 years, loss to follow-up, death, end of the observation period (the latest available data), or study end date (31/12/2025). Participants without a recorded diagnosis of endometriosis or adenomyosis during the study period are administratively censored at the end of follow-up.

Prevalence of endometriosis or adenomyosis among females aged ≥ 12 to ≤ 59 years will be calculated as period prevalence, which summarises the total number of individuals with a recorded diagnosis of interest during a given period, divided by the number of females in the study population during that period. Therefore, period prevalence gives the proportion of individuals who had a recorded diagnosis of

endometriosis or adenomyosis at any time during the specified period. Binomial 95% confidence intervals will be calculated.

Incidence rates and prevalence will be presented overall and stratified by calendar year and age groups.

The denominator for the incidence rate and prevalence estimations will be different depending on the type of data source used, i.e. females receiving care in the contributing hospital system will be the denominator population for hospital-based data sources, while the general population of females will be the denominator population for primary care and registry data sources. As incidence and prevalence estimates may differ substantially across healthcare settings (i.e. primary care and registry data sources versus hospital-based data sources), they should therefore be interpreted within the context of each data source.

An illustration of the calculation of incidence of endometriosis or adenomyosis is shown below in **Figure 3**. Person IDs 1 and 4 will contribute time-at-risk up to the point at which they have a first recorded diagnosis of endometriosis or adenomyosis during the study period. Person IDs 2 and 5 do not have a record of the outcome of interest and so contribute time-at-risk but no incident outcomes. Meanwhile, person ID 3 has a recorded diagnosis of endometriosis or adenomyosis prior to the start of the study period. As a result, this individual will not contribute time-at-risk.

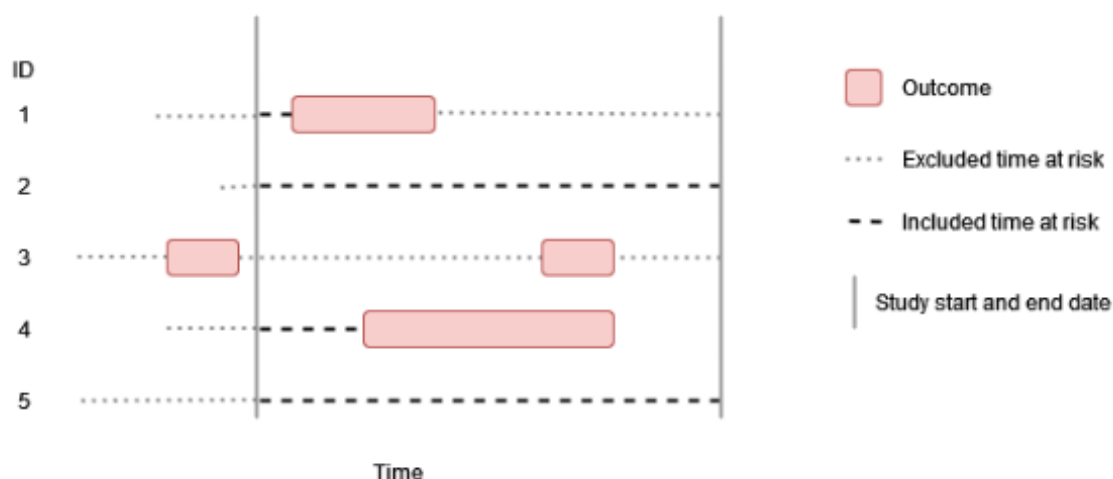


Figure 3. Incidence example.

An illustration of the calculation of period prevalence is shown below in **Figure 4**. Between time $t+2$ and $t+3$, two of the five study participants have a record of the outcome of interest, resulting in a prevalence of 40%. Meanwhile, for the period t to $t+1$, all five also have some observation time during the year, with one of the five study participants having a record of the outcome of interest (i.e. prevalence of 20%).

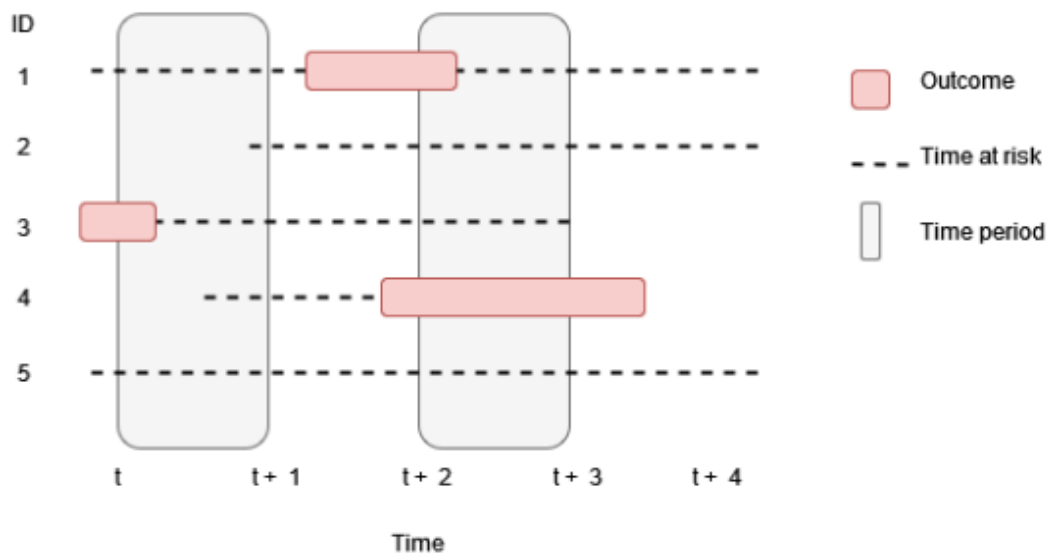


Figure 4. Period prevalence example.

Objectives 2 and 4

The patient-level characterisation of individuals diagnosed with endometriosis or adenomyosis will be estimated based on OMOP CDM mapped data using the *CohortCharacteristics* R package (<https://darwin-eu.github.io/CohortCharacteristics/>), developed by DARWIN EU®.

Individuals diagnosed with endometriosis or adenomyosis will be characterised in terms of age, symptoms related to endometriosis or adenomyosis, most frequently prescribed or dispensed medicines, most frequently recorded procedures, and prespecified treatments for endometriosis or adenomyosis during specific predefined time windows. Categorical variables will be summarised as counts and proportions, while continuous variables will be reported as minimum, q25, median, q75, and maximum.

For the assessment of recorded prespecified symptoms, the total number of individuals, the number of individuals with the prespecified symptom of interest, and the median and interquartile range (IQR) of the number of records per prespecified symptom per individual will be provided per time window. Symptoms recorded before the age of 12 years will not be considered.

The top 10 most frequent recorded procedures will be assessed in hospital-based data sources.

For treatment characterisation, subsequent prescriptions will be combined into a single continuous exposed episode (drug era) if the distance in days between end of the prescription record/dispensing of one prescription and start of the subsequent prescription record/dispensing is ≤ 30 days. This process will be applied iteratively such that multiple consecutive prescription records/dispensings separated by gaps of ≤ 30 days will be combined into a single continuous drug era.

Objective 3

The interval between the first recorded occurrence of any prespecified symptom related to endometriosis or adenomyosis and the first recorded diagnosis of endometriosis or adenomyosis, as well as the interval between the first recorded occurrence of any prespecified symptom related to endometriosis or adenomyosis and the first recorded consultation with gynaecology-related specialty (including reproductive medicine, benign gynaecology, obstetrics, or gynaecological oncology, *if feasible*) will be estimated based on OMOP CDM mapped data using the *CohortCharacteristics* R package (<https://darwin-eu.github.io/CohortCharacteristics/>), developed by DARWIN EU®.

The study population will be restricted to those with at least one recorded prespecified symptom related to endometriosis or adenomyosis prior to the first recorded diagnosis of endometriosis or adenomyosis. For analyses of the interval between the first recorded prespecified symptom and the first consultation with a gynaecology-related specialty, individuals must additionally have at least one recorded consultation with a gynaecology-related specialty after the first recorded symptom and on or before diagnosis. These analyses will be limited to data sources in which specialty information is available and considered suitable for analysis.

Methods to deal with missing data

Analyses will be based on information recorded in the structured source data available in the OMOP CDM. The absence of a recorded diagnosis, symptom, procedure, or treatment does not necessarily indicate the absence of underlying clinical event, as recording practices, healthcare setting, source data availability, and data mapping may vary across data sources.

8.8.4. Output

Output will include a PDF report, including an executive summary, and the following tables and figures.

MOCK Table 1. Study attrition of individuals included in the study population, presented by data source (*objective 1*).

	Data source 1	Data source 2	Data source 3/4/5/6
Qualifying initial records ¹			
Age ≥12 to ≤59 years			
Prior observation of ≥365 days*			
No recorded diagnosis of endometriosis or adenomyosis prior to cohort entry*			

¹Females present in the data source during the study period (01/01/2010 to 31/12/2025 (or the latest available date, if earlier)).

*Only applicable for estimations of incidence rate.

MOCK Table 2. Study attrition of individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, presented by data source (*objectives 2 and 4*).

	Data source 1	Data source 2	Data source 3/4/5/6
Qualifying initial records ¹			
Females aged ≥12 to ≤59 years			
Prior observation of ≥365 days			

¹First recorded diagnosis of endometriosis or adenomyosis during the study period (01/01/2010 to 31/12/2025 (or the latest available date, if earlier)).

MOCK Table 3. Study attrition of individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, presented by data source (*objective 3*).

	Data source 1	Data source 2	Data source 3/4/5/6
Qualifying initial records ¹			
Females aged ≥12 to ≤59 years			

	Data source 1	Data source 2	Data source 3/4/5/6
Prior observation of $\geq 3,650$ days*			
At least one recorded prespecified symptom related to endometriosis or adenomyosis before diagnosis			
At least one recorded consultation with a gynaecology-related specialty after the first recorded symptom**			

¹First recorded diagnosis of endometriosis or adenomyosis during the study period (01/01/2010 to 31/12/2025 (or the latest available date, if earlier)).

*Only applicable to primary care data sources and registries.

**Only applicable for the consultation-specific analysis.

MOCK Table 4. Age at first recorded diagnosis of endometriosis or adenomyosis during 2010–2025 (objective 2).

Variable name	Data source name					
	SUCD	IMASIS	SIDIAP	HI-SPEED	IPCI	CPRD GOLD
Number of individuals						
Age (years), median [q25–q75]						
Age (years), range (min–max)						
Age group, n (%)						
• ≥ 12 to ≤ 17 years						
• ≥ 18 to ≤ 29 years						
• ≥ 30 to ≤ 39 years						
• ≥ 40 to ≤ 49 years						
• ≥ 50 to ≤ 59 years						
Prior observation (days), median (IQR)						
Future observation (days), median (IQR)						

CPRD GOLD = Clinical Practice Research Datalink; HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS = Institut Municipal Assistència Sanitària Information System; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care; SUCD = Semmelweis University Clinical Data; n = number of individuals; IQR = interquartile range.

MOCK Table 5. Prespecified symptoms recorded in individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, assessed across predefined time windows (*objective 2*).

Variable name	Time window	Total individuals (n)	Individuals with ≥1 symptom record, n (%)	Records per individual, median (IQR)
SUCD				
Pelvic or lower abdominal pain				
Dyspareunia				
Urinary tract pain				
Gastrointestinal symptoms				
Excessive menstrual bleeding				
Infertility				
IMASIS/SIDIAP/HI-SPEED/IPCI/CPRD GOLD				
Pelvic or lower abdominal pain, n (%)				
Dyspareunia				
Urinary tract pain				
Gastrointestinal symptoms				
Excessive menstrual bleeding				
Infertility				

CPRD GOLD = Clinical Practice Research Datalink; HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS = Institut Municipal Assistència Sanitària Information System; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care; SUCD = Semmelweis University Clinical Data; n = number of individuals; IQR = interquartile range.

MOCK Table 6. Top 10 medications recorded among individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, assessed in predefined time windows (*objective 2*).

Variable name	Data source name					
	SUCD	IMASIS	SIDIAP	HI-SPEED	IPCI	CPRD GOLD
Time window 1						
Medication 1, n (%)						
Medication 2, n (%)						
Medication 3, n (%)						
Medication 4, n (%)						
Medication 5, n (%)						
Medication 6, n (%)						
Medication 7, n (%)						
Medication 8, n (%)						
Medication 9, n (%)						

Variable name	Data source name					
	SUCD	IMASIS	SIDIAP	HI-SPEED	IPCI	CPRD GOLD
Medication 10, n (%)						
Time window 2/3/4 etc.						
Medication 1, n (%)						
Medication 2, n (%)						
Medication 3, n (%)						
Medication 4, n (%)						
Medication 5, n (%)						
Medication 6, n (%)						
Medication 7, n (%)						
Medication 8, n (%)						
Medication 9, n (%)						
Medication 10, n (%)						

CPRD GOLD = Clinical Practice Research Datalink; HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS = Institut Municipal Assistència Sanitària Information System; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care; SUCD = Semmelweis University Clinical Data; n = number of individuals.

MOCK Table 7. Top 10 procedures recorded among individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, assessed in predefined time windows, in hospital data sources (*objective 2*) (if feasible).

Variable name	Data source name	
	SUCD	IMASIS
Time window 1		
Procedure 1, n (%)		
Procedure 2, n (%)		
Procedure 3, n (%)		
Procedure 4, n (%)		
Procedure 5, n (%)		
Procedure 6, n (%)		
Procedure 7, n (%)		
Procedure 8, n (%)		
Procedure 9, n (%)		
Procedure 10, n (%)		
Time window 2/3/4 etc.		
Procedure 1, n (%)		
Procedure 2, n (%)		
Procedure 3, n (%)		
Procedure 4, n (%)		

Variable name	Data source name	
	SUCD	IMASIS
Procedure 5, n (%)		
Procedure 6, n (%)		
Procedure 7, n (%)		
Procedure 8, n (%)		
Procedure 9, n (%)		
Procedure 10, n (%)		

IMASIS = Institut Municipal Assistència Sanitària Information System; SUCD = Semmelweis University Clinical Data; n = number of individuals.

MOCK Table 8. Time from first recorded prespecified symptom related to endometriosis or adenomyosis to diagnosis and gynaecology-related consultation (*objective 3*).

Data source	Number of individuals with first recorded diagnosis (n)	≥1 prespecified symptom before diagnosis, n (%) ¹	Time from first recorded symptom to diagnosis, median (IQR) days ²	≥1 gynaecology-related consultation after recorded symptom ³	Time from first recorded symptom to first recorded gynaecology-related consultation, median (IQR) days ⁴
SUCD					
IMASIS					
SIDIAP					
HI-SPEED					
IPCI					
CPRD GOLD					

CPRD GOLD = Clinical Practice Research Datalink; HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS = Institut Municipal Assistència Sanitària Information System; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care; SUCD = Semmelweis University Clinical Data; n = number of individuals; IQR = interquartile range.

¹The percentage denominator is the number of individuals with a first recorded diagnosis who meet the applicable age and prior data visibility criteria.

²Calculated among individuals with at least one prespecified symptom recorded before their first recorded diagnosis.

³The percentage denominator is the number of individuals with at least one prespecified symptom recorded before diagnosis in data sources where gynaecology-related specialty information is available and considered suitable for analysis.

⁴Calculated only among individuals with a recorded gynaecology-related consultation occurring after the first recorded prespecified symptom.

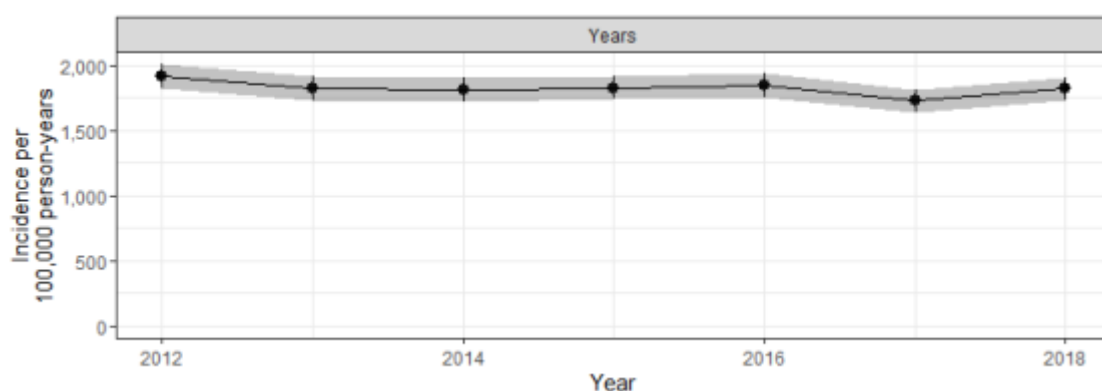
MOCK Table 9. Treatments and surgical procedures recorded among individuals with a first recorded diagnosis of endometriosis or adenomyosis during 2010–2025, assessed in predefined time windows (*objective 4*).

Treatment class	Ingredient/substance*	Data source name					
		SUCD	IMASIS	SIDIAP	HI-SPEED	IPCI	CPRD GOLD
Time window 1							
Combined hormonal contraceptives	Overall, n (%)						
	Megestrol, n (%)						
	Medroxyprogesterone, n (%)						
	Mestranol, n (%)						
Progestogens	Overall						
	Norethindrone, n (%)						
	Dienogest, n (%)						
	Medroxyprogesterone, n (%)						
GnRH agonists	Overall, n (%)						
	Goserelin, n (%)						
	Histrelin, n (%)						
	Buserelin, n (%)						
GnRH antagonists	Overall, n (%)						
	Elagolix, n (%)						
	Linzagolix, n (%)						
	Relugolix, n (%)						
Aromatase inhibitors	Overall, n (%)						
	Letrozole, n (%)						
	Anastrozole, n (%)						

Treatment class	Ingredient/substance*	Data source name					
		SUCD	IMASIS	SIDIAP	HI-SPEED	IPCI	CPRD GOLD
	Exemestane, n (%)						
Surgery, n (%)							
Time window 2/3/4 etc.							
Combined hormonal contraceptives							
Progestogens							
GnRH agonists							
GnRH antagonists							
Aromatase inhibitors							
Surgery							

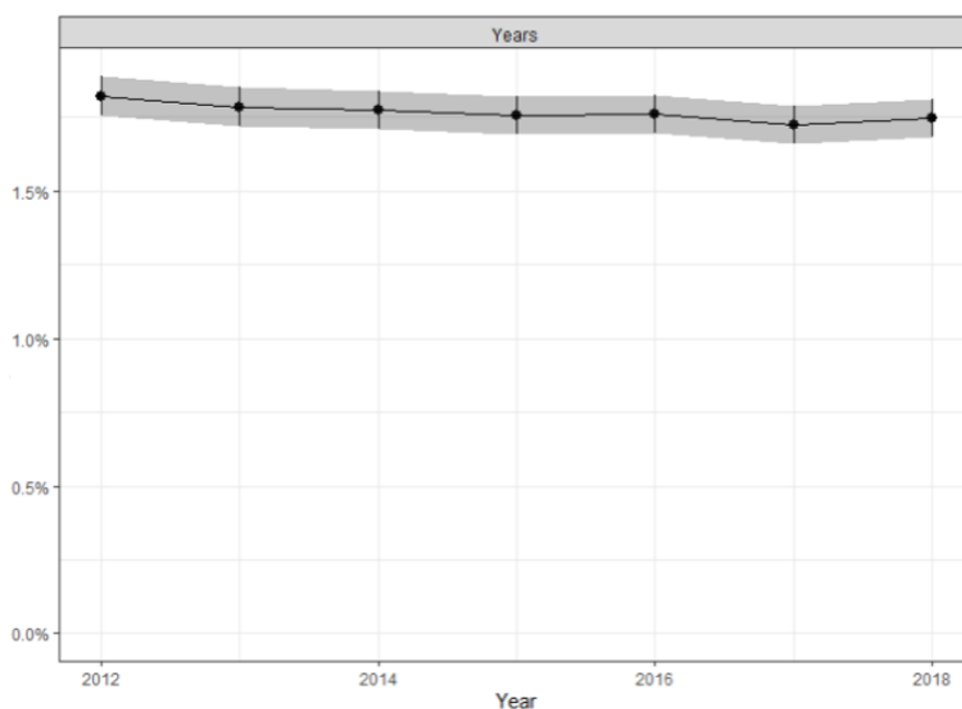
CPRD GOLD = Clinical Practice Research Datalink; HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS = Institut Municipal Assistència Sanitària Information System; IPCI = Integrated Primary Care Information; SIDIAP = The Information System for Research in Primary Care; SUCD = Semmelweis University Clinical Data; n = number of individuals.

*=For each treatment class, results will be presented both at the class (overall) level and ingredient/substance level. Only selected examples are shown in this mock table.



MOCK Figure 1. Annual incidence rates of endometriosis or adenomyosis diagnosis during the study period (objective 1).

Incident rates will be provided overall and stratified by age group.



MOCK Figure 2. Annual prevalence of endometriosis or adenomyosis diagnosis during the study period (objective 1).

Prevalence graphs will be provided for endometriosis and adenomyosis separately, and for both conditions combined. Prevalence estimates will be provided overall and stratified by age group.

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

8.9. Evidence synthesis

Results from analyses described in [Section 8.8](#) will be presented separately for each data source. No meta-analysis of results will be conducted.

9. STRENGTHS AND LIMITATIONS

Strengths

This study leverages a large and diverse set of real-world healthcare data sources from Hungary, Spain, Sweden, the Netherlands, and the United Kingdom, including primary care data sources, electronic health records, hospital-based data, and population-based registry data. The broad geographic coverage and diversity of healthcare settings will enable a comprehensive assessment of the epidemiology of endometriosis and adenomyosis across Europe. Use of the OMOP Common Data Model (CDM) ensures standardised data structuring, harmonised variables, and consistent cohort definitions across data sources. This facilitates reproducible analyses and enables estimation of incidence and prevalence, as well as characterisation of demographic and clinical characteristics, symptoms, diagnostic delay, and recorded treatments among females with recorded endometriosis or adenomyosis.

Limitations

The study will be informed by routinely collected healthcare data, and therefore, several limitations should be considered.

Diagnoses of endometriosis and adenomyosis may be subject to misclassification due to differences in coding practices, healthcare utilisation, and disease recognition across healthcare settings and data sources. The first recorded diagnosis in a data source may not represent the individual's first clinical diagnosis if an earlier diagnosis occurred before entry into the data source or in a healthcare setting not captured by that data source. Symptoms related to endometriosis or adenomyosis may be incompletely captured in structured healthcare data, particularly when documented in free text or not recorded during healthcare encounters, which may result in underestimation of symptom frequencies and the recorded diagnostic interval. Similarly, treatment information may be incompletely captured depending on the data source and care setting.

The assessment of diagnostic delay will be based on the first recorded symptom rather than the true onset of symptoms and should therefore be interpreted as reflecting the time from first recorded prespecified symptom to first recorded diagnosis in healthcare records. Accordingly, the analysis will describe the interval between the first recorded prespecified symptom and the first recorded diagnosis among diagnosed females and should be interpreted as a measure of time to recorded diagnosis rather than true clinical diagnostic delay. In addition, some prespecified symptoms of endometriosis and adenomyosis are non-specific and may be incompletely or inconsistently captured in structured healthcare data. Such symptoms may also be attributable to conditions other than endometriosis or adenomyosis, which should be considered when interpreting symptom frequencies and time to diagnosis estimates. Furthermore, recorded diagnostic interval estimates as proxy for diagnosis delay will depend on the available lookback period and duration of patient observation before diagnosis. Consequently, this measure may underestimate or misclassify true clinical diagnosis delay. Requiring at least 10 years of prior data visibility for individuals from primary care data sources and registries may introduce selection by restricting the analysis to individuals with long-term and stable observation in the data source. In addition, restricting the analysis to diagnosed individuals with at least one prespecified symptom recorded before diagnosis excludes diagnosed individuals whose symptoms were not captured in structured data.

The first recorded consultation with a gynaecology-related specialty may not necessarily relate to the investigation or management of endometriosis or adenomyosis. Specialty information may also be incomplete or recorded differently across data sources. Analyses involving gynaecology-related consultations will therefore be limited to data sources in which the relevant specialty information is available and considered suitable for analysis. Results from this analysis should be interpreted as a complementary measure of the recorded diagnostic pathway rather than a direct measure of diagnostic delay.

Although all participating data sources are mapped to the OMOP Common Data Model and common analytical procedures will be applied, differences in source populations, healthcare systems, coding practices, and data availability may affect the comparability of results across data sources. Differences in healthcare setting (e.g. primary care, hospital-based, or registry data) may additionally influence the recording and completeness of diagnoses, symptoms, diagnostic procedures, specialist consultations, and treatments. Finally, the findings will reflect the populations represented in the participating data sources from Hungary, Spain, Sweden, the Netherlands, and the United Kingdom and may not be generalisable to other populations or healthcare settings.

10. REFERENCES

1. Wang MH, Chen JH, Qi XY, Li ZX, Huang Y. *Global prevalence of adenomyosis and endometriosis: a systematic review and meta-analysis*. Reproductive Biology and Endocrinology. 2025. **23**(1): p.148.
2. Ochoa Bernal, M.A. and A.T. Fazleabas, *The Known, the Unknown and the Future of the Pathophysiology of Endometriosis*. International Journal of Molecular Sciences, 2024. **25**(11): p. 5815.
3. Allaire, C., M.A. Bedaiwy, and P.J. Yong, *Diagnosis and management of endometriosis*. Canadian Medical Association Journal, 2023. **195**(10): p. E363-E371.
4. Breton Z, Gouesbet S, Indersie E, Gabillet M, Tran VT, Aflak N, Borghese B, Petit E, Roman H, Millepied AC, Nève De Mevergnies M. *Endometriosis diagnostic delay and its correlates: results from the ComPaRe-Endometriosis cohort*. Journal of Women's Health. 2026. **35**(2): p. 172-88.
5. Medina-Perucha L, Pistillo A, Raventós B, Jacques-Aviñó C, Munrós-Feliu J, Martínez-Bueno C, Valls-Llobet C, Carmona F, López-Jiménez T, Pujolar-Díaz G, Flo Arcas E. *Endometriosis prevalence and incidence trends in a large population-based study in Catalonia (Spain) from 2009 to 2018*. Women's Health. 2022. **18**:17455057221130566.

11. ANNEXES

ANNEX I. Description of data sources

Hungary: Semmelweis University Clinical Data (SUCD)

#	Section	Description
1	Data source identification and country	SUCD (Semmelweis University Clinical Data) Budapest, Hungary
2	Data partner information section	Semmelweis University (SU) -
3	Coverage and timespan	Data collection since: 2010 Extent: Regional. The general catchment area of SU is the central region of the country, Budapest city and Pest county, although patients can be referred from anywhere in Hungary. The total population of Budapest and Pest county is approximately 4,200,000 people. The total population of Hungary is around 9,500,000.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data, and other (specify). diagnostic data (laboratory tests, radiology, pathology)
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems, and Registries. Data is extracted directly from the source database. From there, the data entry in the system is heavily controlled and validated on the user interface before being made available for further research.
6	General representativeness	SU captures information on patients who are covered by the public health insurance system. This covers all Hungarian citizens, and therefore the database should mirror the source population well. Although, besides Semmelweis University Clinics, there are multiple hospitals in the region, and data on visits in other hospitals is not represented in the database. Therefore the patient population is not directly representative of the general population.
7	Data content / source coding	Regarding SU's source data, procedures and diagnoses are coded in SNOMED, measurements are coded in LOINC, and drugs are stored in RxNorm and ATC.
8	Data Harmonisation	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 have a unique identifier (SSN).
9	Quality control (data source specific)	The clinical database is the source database and therefore it has to be treated as a trusted database. Data entry in the systems is heavily controlled by validation on the user interface, and there are large number of rules that controls the data on the insurer's side that has to be corrected in the system by the users to be able to close the encounters. OMOP mapping is done in the framework by EHDEN recognized partners under quality check by the EHDEN society.
10	Linkage	No known linkages.
11	Vital status	Source for vital status unknown.
12	Limitations	Medication prescribed in secondary care is fully present in our database, but medication given in the hospital is rarely documented. General limitations for the data type applicable. General practitioner data is not present in our database; therefore, this part of the patient journey is not represented.
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/1000000184 Website: https://www.semmelweis.hu

Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)

#	Section	Description
1	Data source identification and country	IMASIS (Institut Municipal Assistència Sanitària Information System) Catalunya, Spain
2	Data partner information section	Parc de Salut Mar Barcelona (PSMAR) Direcció de Control de Gestió (Management and Control Department)
3	Coverage and timespan	Data collection since: 1990 Extent: Regional. The primary catchment area covers the Barcelona city area, in Catalonia (Spain). However, all recorded visits are included, which may also include patients residing outside this area.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The available information includes diagnoses, procedures, drug administrations, dispensations and prescriptions, laboratory tests, visit types, cancer data, and vaccinations. These data are linked to primary care centers, which currently contribute information on diagnoses, and treatments, as well as to mortality data.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. The data collection process begins at the point of care, when a patient is attended in any facility within our hospital network or associated primary care centers. Clinical information is recorded directly into the electronic health records (EHRs) and subsequently complemented with data from external sources, including the regional vaccine registry, the cancer registry, the national mortality registry, and outpatient treatment records. All information is consolidated into a relational database using standardized formats and subjected to multiple quality control procedures to ensure completeness and consistency. The resulting datasets are then extracted, transformed, and loaded (ETL) into an OMOP Common Data Model (CDM) database on a quarterly to semi-annual basis.
6	General representativeness	The data source includes patients across all socioeconomic levels, ages, and genders, suggesting that the dataset is representative of the population of Barcelona requiring secondary (hospital-based) healthcare at any given time.
7	Data content / source coding	Health outcomes, diagnoses, and procedures are recorded using ICD-9-CM and ICD-10-CM classification. Laboratory tests are recorded in the source data using LOINC terminology and drugs using the national codes classification. Cancer registry is based on ICD-O-3 classification. Additionally, a set of demographics data are available such as age and sex.
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 cannot be registered under different ID numbers. They are registered under the same ID, even when changing practice of re-registration or admission.
9	Quality control (data source specific)	Several quality control procedures have been implemented during the creation and update of IMASIS. The process includes a series of filters that exclude any data whose format or content do not match the specifications of the target variables. Some examples of these actions include the following. - Records containing codes that do not belong to the standardized vocabularies used for encoding are reviewed within their respective domains and, if necessary, discarded. - Records with implausible values, such as dates preceding a patient's birthdate, are identified and corrected or removed. - Before each IMASIS update, the number of records in the source files is verified. Record counts from the new IMASIS instance are compared with both the source files and previous IMASIS versions to detect potential inconsistencies or anomalies in the source data. - In addition, if the Data Quality Dashboard (DQD) package identifies quality issues in the OMOP instance, for example diagnoses recorded for patients with an implausible gender, these discrepancies are reviewed and corrected in the source data.

#	Section	Description
10	Linkage	Information of diagnosis and treatments received in outpatient settings and in primary care as well as national mortality registry, cancer registry and regional vaccination registry.
11	Vital status	The hospital's death registry records the age at death for each patient when the event occurs within the hospital. In addition, deaths occurring outside the hospital are captured through linkage with the national mortality registry; however, in these cases, the specific cause of death is not currently available.
12	Limitations	No database-specific limitations are present but information coming from primary care are not complete. General limitations related to Real World Data can be considered.
13	Main references	Mayer MA, Furlong LI, Torre P, Planas I, Cots F, Izquierdo E, Portabella J, Rovira J, Gutierrez-Sacristan A, Sanz F "Reuse of EHRs to Support Clinical Research in a Hospital of Reference." Studies in health technology and informatics (2015): 25991136
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104316 Website: https://www.parcdesalutmar.cat/en/

Spain: 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, 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 Català 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.

#	Section	Description
		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, Bolívar B, Aragón M, Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

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

#	Section	Description
1	Data source identification and country	HI-SPEED (Health Impact - Swedish Population Evidence Enabling Data-linkage) Sweden
2	Data partner information section	Swedish Medical Products Agency - Gothenburg University (SMPA-GU) Pharmacoepidemiology and Analysis Department (FeA) - School of Public Health and Community Medicine, Institute of Medicine
3	Coverage and timespan	Data collection since: 2015 Extent: Nation-wide. The catchment area includes the whole of Sweden, covering the full population of approximately 11.7 million.
4	Healthcare setting / type of data	Primary care – General Practitioner, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Primary care (GPs) is available only for the 2 largest regions (~40% of national population) The following data elements are collected: Socio-demographics, dispensed drug prescriptions, cause of death, diagnoses and procedures from secondary (specialist) care and inpatient visits or clinical events, as well as from primary care visits (40%pop only).
5	Data collection process	Registries. The data is acquired from the relevant Swedish national and regional registries, only once all

#	Section	Description
		legislative, GDPR and ethical approvals have been granted. Therefore only relevant data is passed on, which will then be entered and processed by the study team. The data are updated several times annually.
6	General representativeness	The coverage includes all patients of all sociodemographic characteristics. Therefore it should mirror the source population to a very good extent.
7	Data content / source coding	Medicines are coded with ATC and NPLID (National Product ID), ICD10-SE is used for diagnoses and the Swedish procedure coding system (KVA) is used for clinical procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients have a uniquely id across datasets.
9	Quality control (data source specific)	The source data are obtained from the relevant Swedish National and Regional Registers. The registers perform some regular quality controls on their data. After receiving the data, we perform additional checks and cleaning. We also run regular quality checks on the data we manage.
10	Linkage	Data on specialist care is acquired from the National Patient Register, mortality information is provided by the Cause-Of-Death Registry. Drug data is provided by the Patient Drug Register. And similarly for other registers. Data are linked very accurately using the national personal ID number, and pseudonymized before delivery to HI-SPEED. All data are updated 2-4 times per year.
11	Vital status	Data on death and underlying + contributing causes-of-death are extracted from the Cause-of-Death registry (i.e. based on death certificates).
12	Limitations	General limitations for the data type applicable. This is a research project where all studies require ethics approval. Data collection since: 2015 for most data, except prescribed drug register (from 2018), and some COVID-related data (tests, vaccination) from 2020 Primary care is only available for a subset.
13	Main references	Nyberg F, Franzén S, Lindh M, Vanfleteren L, Hammar N, Wettermark B, Sundström J, Santosa A, Björck S, Gisslén M "Swedish Covid-19 Investigation for Future Insights - A Population Epidemiology Approach Using Register Linkage (SCIFI-PEARL)." Clinical epidemiology (2021): 34354377
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/node/4463/ Website: https://www.gu.se/en/research/scifi-pearl

The Netherlands: Integrated Primary Care Information (IPCI)

#	Section	Description
1	Data source identification and country	IPCI (Integrated Primary Care Information) the Netherlands
2	Data partner information section	Erasmus University Medical Center (EMC) Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards. However, it mainly covers the central part of the country, including the most densely populated area (the 'Randstad') and non-urban areas. IPCI contains information on all patients registered with GPs responsible for non-emergency care and referrals. A patient is registered at birth or at first encounter with the GP.

#	Section	Description
4	Healthcare setting / type of data	Primary care – General Practitioner. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. The patientdossiers are collected by Erasmus MC data managers and combined in one harmonized database. Several checks are done on this database to ensure correct data processing. Persons can have dossiers at multiple GPs.
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content / source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data 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 can be registered under different IDs, but since a patient can only be registered at one GP at a time, the observations periods will not overlap.
9	Quality control (data source specific)	Prior to each data release, extensive quality control steps are performed, e.g., comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g. number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical data, e.g. distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Capture of oral contraceptives in IPCI is incomplete. The first prescription is recorded within the patient's medical file but not the subsequent prescriptions. If you want to use IPCI data for research questions on oral contraceptives, you can do that but then you need to select those GPs that use MEDICOM as software system (where you have prescribing and dispensing). However, of hormone substitution therapy, the information on prescribing is COMPLETE. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.

#	Section	Description
13	Main references	de Ridder MAJ, de Wilde M, de Ben C, Leyba AR, Mosseveld BMT, Verhamme KMC, van der Lei J, Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182144
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

The United Kingdom: 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

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

ANNEX II. Fitness for use assessment

Data source justification for inclusion and key characteristics

Hungary: Semmelweis University Clinical Data (SUCD)

SUCD will be included in this study because it is a hospital data source that provides relevant information on endometriosis or adenomyosis in the female population aged ≥ 12 to ≤ 59 years.

Based on a preliminary feasibility assessment, the expected number of person-counts for endometriosis in SUCD will be 17,800.

Moreover, data availability and follow-up in SUCD is sufficient, as data availability starts in 2010, and the date of the most recent data extraction (as of August 2026) is 06/2026, which aligns with the study period. The median follow-up of the first observation period in SUCD is 237 days.

There are no study specific limitations present in SUCD.

Lastly, Institutional Review Board (IRB) approval for SUCD is estimated to take 1 – 3 months, which makes the execution of this study feasible within the current study timelines.

Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)

IMASIS will be included in this study because it is a hospital data source that provides relevant information on endometriosis or adenomyosis in the female population aged ≥ 12 to ≤ 59 years.

Based on a preliminary feasibility assessment, the expected number of person-counts for endometriosis in IMASIS will be 2,600.

Moreover, data availability and follow-up in IMASIS is sufficient, as data availability starts in 1990, and the date of the most recent data extraction (as of August 2026) is 01/2026, which aligns with the study period. The median follow-up of the first observation period in IMASIS is 1,310 days.

There are no study specific limitations present in IMASIS.

Lastly, IRB approval for IMASIS is estimated to take 1 – 3 months, 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 endometriosis or adenomyosis in the female population aged ≥ 12 to ≤ 59 years.

Based on a preliminary feasibility assessment, the expected number of person-counts for endometriosis in SIDIAP will be 45,400.

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 12/2024, which aligns with the study period. The median follow-up of the first observation period is 5,760 days.

There are no study specific limitations present in SIDIAP.

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

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

HI-SPEED will be included in this study because it is a registry data source that provides relevant information on endometriosis or adenomyosis in the female population aged ≥ 12 to ≤ 59 years.

Based on a preliminary feasibility assessment, the expected number of person-counts for endometriosis in HI-SPEED will be 54,100.

Moreover, data availability and follow-up in HI-SPEED is sufficient, as data availability starts in 2015, and the date of the most recent data extraction (as of August 2026) is 12/2025, which aligns with the study period. The median follow-up of the first observation period in HI-SPEED is 3,890 days.

There are some study specific limitations present in HI-SPEED, namely: due to data availability, the study period for HI-SPEED will be from 2015 onwards.

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

The Netherlands: Integrated Primary Care Information (IPCI)

IPCI will be included in this study because it is a primary care data source that provides relevant information on endometriosis or adenomyosis in the female population aged ≥ 12 to ≤ 59 years.

Based on a preliminary feasibility assessment, the expected number of person-counts for endometriosis in IPCI will be 13,600.

Moreover, data availability and follow-up in IPCI is sufficient, as data availability starts in 2006, and the date of the most recent data extraction (as of August 2026) is 12/2025, which aligns with the study period. The median follow-up of the first observation period is 1,830 days.

There are no study specific limitations present in IPCI.

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

The United Kingdom: 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 endometriosis or adenomyosis in the female population aged ≥ 12 to ≤ 59 years.

Based on a preliminary feasibility assessment, the expected number of person-counts for endometriosis in CPRD GOLD will be 53,200.

Moreover, data availability and follow-up in CPRD GOLD is sufficient, as data availability starts 1987, and the date of the most recent data extraction (as of August 2026) is 11/2025, which aligns with the study period. The median follow-up of the first observation period is 2,160 days.

There are no study specific limitations present in CPRD GOLD.

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

Table S1. Fitness-for-use assessment of data sources.

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant	Suggested extensiveness assessment	Suggested assessment of other quality dimensions	Suggested substantiation by documentation
Study population	Females aged ≥ 12 to ≤ 59 years (<i>all objectives</i>)	Sex and year of birth	High	100% of included individuals will have sex and year of birth available	N/A	N/A
	Recorded diagnosis of endometriosis or adenomyosis during the study period (<i>objectives 2, 3, 4</i>)	SNOMED concept IDs for endometriosis or adenomyosis	High	100% of included individuals will have a diagnosis of endometriosis or adenomyosis	N/A	N/A
Treatment/ exposure	N/A	N/A	N/A	N/A	N/A	N/A
Comparator group (not relevant)	N/A	N/A	N/A	N/A	N/A	N/A
Outcomes	Recorded diagnosis of endometriosis or adenomyosis (<i>objective 1</i>)	SNOMED concept IDs for endometriosis or adenomyosis	Medium	N/A (to be assessed on a research question basis)	N/A	N/A
	Characteristics of individuals with endometriosis or adenomyosis (<i>objective 2</i>)	Year of birth and concept IDs for prespecified symptoms, procedures, and comedication	Medium	N/A (to be assessed on a research question basis)	N/A	N/A
	Time from first recorded symptom to diagnosis or gynaecology-related consultation (<i>objective 3</i>)	Concept IDs for prespecified symptoms and gynaecology-related consultation	High	100% of included individuals will have at least one recorded prespecified symptom of interest prior to diagnosis. For analyses of time from first symptom to first		

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant	Suggested extensiveness assessment	Suggested assessment of other quality dimensions	Suggested substantiation by documentation
				gynaecology-related consultation, 100% of included individuals will also have a recorded consultation. These analyses will only be conducted in data sources where gynaecology-related specialty information is available and considered suitable for analysis.		
	Treatments (<i>objective 4</i>)	Prescription of prespecified treatments (RxNorm code) or surgical procedure record	Medium	N/A (to be assessed on a research question basis)	N/A	N/A
Covariates	Age group strata (<i>objective 1</i>)	Year of birth	Medium	100% of included individuals will have year of birth available	N/A	N/A
Follow-up time (not relevant)	N/A	N/A	N/A	N/A	N/A	N/A

N/A = Not applicable, RxNorm = Medical prescription normalised, SNOMED = Systematized Nomenclature of Medicine.

EMA Data Quality Framework for EU medicines regulation: application to Real-World Data for more information

(https://www.ema.europa.eu/system/files/documents/other/data-quality-framework-eu-medicines-regulation-application-real-world-data_en.pdf).

ANNEX III. Operational and reporting considerations

DATA MANAGEMENT

Data management

All data sources have previously mapped their data to the OMOP common data model. This enables the use of standardised analytics and using DARWIN EU® tools across the network, since the structure of the data and the terminology system 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, participants from various EU member states will process personal data from individuals that is collected in national/regional electronic health record data sources. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

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

The output files are stored in the DARWIN EU® 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 *DrugExposureDiagnostics* (<https://cran.r-project.org/web/packages/DrugExposureDiagnostics/index.html>) R packages will be run, if needed, to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error. The *DrugExposureDiagnostics* package evaluates ingredient-specific attributes and patterns in drug exposure records.

The study code will be based on DARWIN EU® R packages: *IncidencePrevalence* to estimate incidence and prevalence and *CohortCharacteristics* to characterise study population. 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. Lists with concept definitions

List of conditions and procedure concepts definition can be provided in the table below:

Table S2. Preliminary list of conditions or procedure definitions.

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Endometriosis	Endometriosis (clinical)	433527	no	SNOMED
Endometriosis	Endometriosis	4244991	no	SNOMED
Endometriosis	Endometriotic cyst	4014020	no	SNOMED
Adenomyosis	Uterine adenomyosis	37205049	no	SNOMED
Adenomyosis	Abnormal uterine bleeding due to adenomyosis	37209637	no	SNOMED
Adenomyosis	Adenomyosis of fallopian tube	1075237	no	SNOMED
Pelvic or lower abdominal pain	Pain in pelvis	4147829	no	SNOMED
Pelvic or lower abdominal pain	Lower abdominal pain	4182562	no	SNOMED
Dyspareunia	Dyspareunia	439080	no	SNOMED
Urinary tract pain	Urinary tract pain	4082798	no	SNOMED
Gastrointestinal symptoms	Bowel dysfunction	4341223	no	SNOMED
Gastrointestinal symptoms	Diarrhea	196523	no	SNOMED
Gastrointestinal symptoms	Constipation	75860	no	SNOMED
Gastrointestinal symptoms	Abdominal bloating	4023572	no	SNOMED
Gastrointestinal symptoms	Pain associated with defecation	4025947	no	SNOMED
Excessive menstrual bleeding	Mid-cycle bleeding	4129545	no	SNOMED
Excessive menstrual bleeding	Menorrhagia	4302555	no	SNOMED
Excessive menstrual bleeding	Prolonged periods	4088438	no	SNOMED
Excessive menstrual bleeding	Abnormal uterine bleeding	45757810	no	SNOMED
Infertility	Female infertility	201909	no	SNOMED
Infertility	Decreased fertility	4066416	no	SNOMED
Infertility	Decreased fertility in male	37167145	yes	SNOMED
Surgery	Surgical procedure	4301351	no	SNOMED

List of medicines concepts definition can be provided in the table below:

Table S3. Preliminary list of medicines definitions.

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendant s
megestrol	megestrol	combined hormonal contraceptives	G03AA; G03AB	1300978	yes
medroxyprogesterone	medroxyprogesterone	combined hormonal contraceptives	G03AA; G03AB	1500211	yes
mestranol	mestranol	combined hormonal contraceptives	G03AA; G03AB	1503184	yes
drospirenone	drospirenone	combined hormonal contraceptives	G03AA; G03AB	1512674	yes
norgestimate	norgestimate	combined hormonal contraceptives	G03AA; G03AB	1515774	yes
norelgestromin	norelgestromin	combined hormonal contraceptives	G03AA; G03AB	1518198	yes
norethindrone	norethindrone	combined hormonal contraceptives	G03AA; G03AB	1521369	yes
norgestrel	norgestrel	combined hormonal contraceptives	G03AA; G03AB	1521592	yes
estradiol	estradiol	combined hormonal contraceptives	G03AA; G03AB	1548195	yes
ethinyl estradiol	ethinyl estradiol	combined hormonal contraceptives	G03AA; G03AB	1549786	yes
desogestrel	desogestrel	combined hormonal contraceptives	G03AA; G03AB	1588000	yes
levonorgestrel	levonorgestrel	combined hormonal contraceptives	G03AA; G03AB	1589505	yes
ethynodiol	ethynodiol	combined hormonal contraceptives	G03AA; G03AB	1595461	yes
nomegestrol	nomegestrol	combined hormonal contraceptives	G03AA; G03AB	19015533	yes
chlormadinone	chlormadinone	combined hormonal contraceptives	G03AA; G03AB	19026733	yes
dienogest	dienogest	combined hormonal contraceptives	G03AA; G03AB	19054876	yes
lynestrenol	lynestrenol	combined hormonal contraceptives	G03AA; G03AB	19092358	yes
gestodene	gestodene	combined hormonal contraceptives	G03AA; G03AB	19097684	yes
estetrol	estetrol	combined hormonal contraceptives	G03AA; G03AB	36026931	yes
QUINGESTANOL	QUINGESTANOL	combined hormonal contraceptives	G03AA; G03AB	36849489	yes

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
NORMETHANDRONE	NORMETHANDRONE	Progestogens	G03DC	36853507	yes
QUINGESTANOL	QUINGESTANOL	Progestogens	G03AC	36849489	yes
17-alpha-hydroxyprogesterone	17-alpha-hydroxyprogesterone	Progestogens	G03DA	19077143	yes
allylestrenol	allylestrenol	Progestogens	G03DC	19068397	yes
chlormadinone	chlormadinone	Progestogens	G03DB	19026733	yes
demegestone	demegestone	Progestogens	G03DB	19052689	yes
desogestrel	desogestrel	Progestogens	G03AC	1588000	yes
dienogest	dienogest	Progestogens	G03DB	19054876	yes
drospirenone	drospirenone	Progestogens	G03AC	1512674	yes
dydrogesterone	dydrogesterone	Progestogens	G03DB	19040060	yes
ethisterone	ethisterone	Progestogens	G03DC	19050090	yes
ethynodiol	ethynodiol	Progestogens	G03DC	1595461	yes
etonogestrel	etonogestrel	Progestogens	G03AC	1519936	yes
gestonorone caproate	gestonorone caproate	Progestogens	G03DA	19059487	yes
levonorgestrel	levonorgestrel	Progestogens	G03AC	1589505	yes
lynestrenol	lynestrenol	Progestogens	G03DC	19092358	yes
medrogestone	medrogestone	Progestogens	G03DB	19000207	yes
medroxyprogesterone	medroxyprogesterone	Progestogens	G03DA	1500211	yes
megestrol	megestrol	Progestogens	G03DB	1300978	yes
nomegestrol	nomegestrol	Progestogens	G03DB	19015533	yes
norethindrone	norethindrone	Progestogens	G03AC	1521369	yes
norgestrienone	norgestrienone	Progestogens	G03AC	19128258	yes
progesterone	progesterone	Progestogens	G03DA	1552310	yes
gonadorelin	gonadorelin	GnRH agonists	H01CA01	19089810	yes
nafarelin	nafarelin	GnRH agonists	H01CA02	1507558	yes

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
goserelin	goserelin	GnRH agonists	L02AE	1366310	yes
histrelin	histrelin	GnRH agonists	L02AE	1366773	yes
buserelin	buserelin	GnRH agonists	L02AE	19033280	yes
triptorelin	triptorelin	GnRH agonists	L02AE	1343039	yes
leuprolide	leuprolide	GnRH agonists	L02AE	1351541	yes
cetrorelix	cetrorelix	GnRH antagonists	H01CC	1503983	yes
ganirelix	ganirelix	GnRH antagonists	H01CC	1536743	yes
elagolix	elagolix	GnRH antagonists	H01CC	1560104	yes
linzagolix	linzagolix	GnRH antagonists	H01CC	741911	yes
relugolix	relugolix	GnRH antagonists	H01CC	739471	yes
letrozole	letrozole	Aromatase inhibitors	L02BG	1315946	yes
anastrozole	anastrozole	Aromatase inhibitors	L02BG	1348265	yes
exemestane	exemestane	Aromatase inhibitors	L02BG	1398399	yes
aminoglutethimide	aminoglutethimide	Aromatase inhibitors	L02BG	1503057	yes
formestane	formestane	Aromatase inhibitors	L02BG	19020079	yes
VOROZOLE	VOROZOLE	Aromatase inhibitors	L02BG	36849540	yes

Combined hormonal contraceptive concept sets were informed by WHO ATC codes G03AA and G03AB. Ingredient concepts were included and descendants were used to capture corresponding fixed-dose combination and sequential products (e.g. ethinyl estradiol/progestogen, estradiol/progestogen, and estetrol/progestogen combinations).

List of gynaecology-related specialty concepts definition can be provided in the table below:

Table S4. Preliminary list of gynaecology-related specialty definitions.

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Gynaecology-related specialty	Obstetrics / Gynaecology	38004461	no	Medicare Specialty

ANNEX V. Illustrative examples of follow-up and censoring for incidence and prevalence analyses

An example of entry and exit into the denominator population to estimate incidence rates is shown in **Figure S1**. In this example, person ID 1 already has sufficient prior history before the study start date, and the observation period ends after the study end date, so this individual will contribute during the complete study period. Person IDs 2 and 4 enter the study only when they have sufficient prior history. Person ID 3 leaves when exiting the data source (the end of the observation period). Lastly, person ID 5 has two observation periods in the data source. The first period contributes time from the study start until the end of the observation period. The second starts contributing time again once sufficient prior history is reached and exits at the study end date.

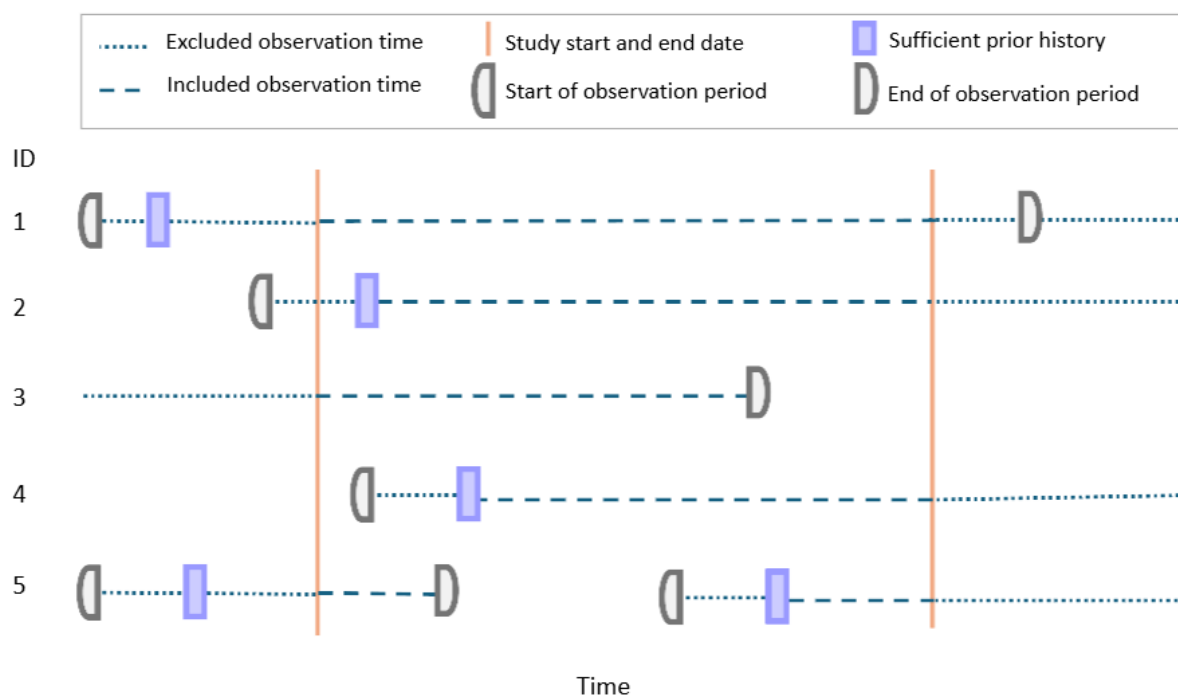


Figure S1. Included observation time for the denominator population for incidence rates.

An illustrative example of entry and exit into the denominator population to estimate prevalence is shown in **Figure S2**. In this example, the observation period of person IDs 1 and 2 starts before the study start date and the observation period ends after the study end date, so these individuals will contribute during the complete study period. Person ID 3 leaves when exiting the data source (the end of the observation period). Person ID 4 enters the study when their observation period starts. Lastly, person ID 5 has two observation periods in the data source. The first period contributes time from the study start until the end of the observation period, the second starts contributing time again once the observation period starts and exits at the study end date.

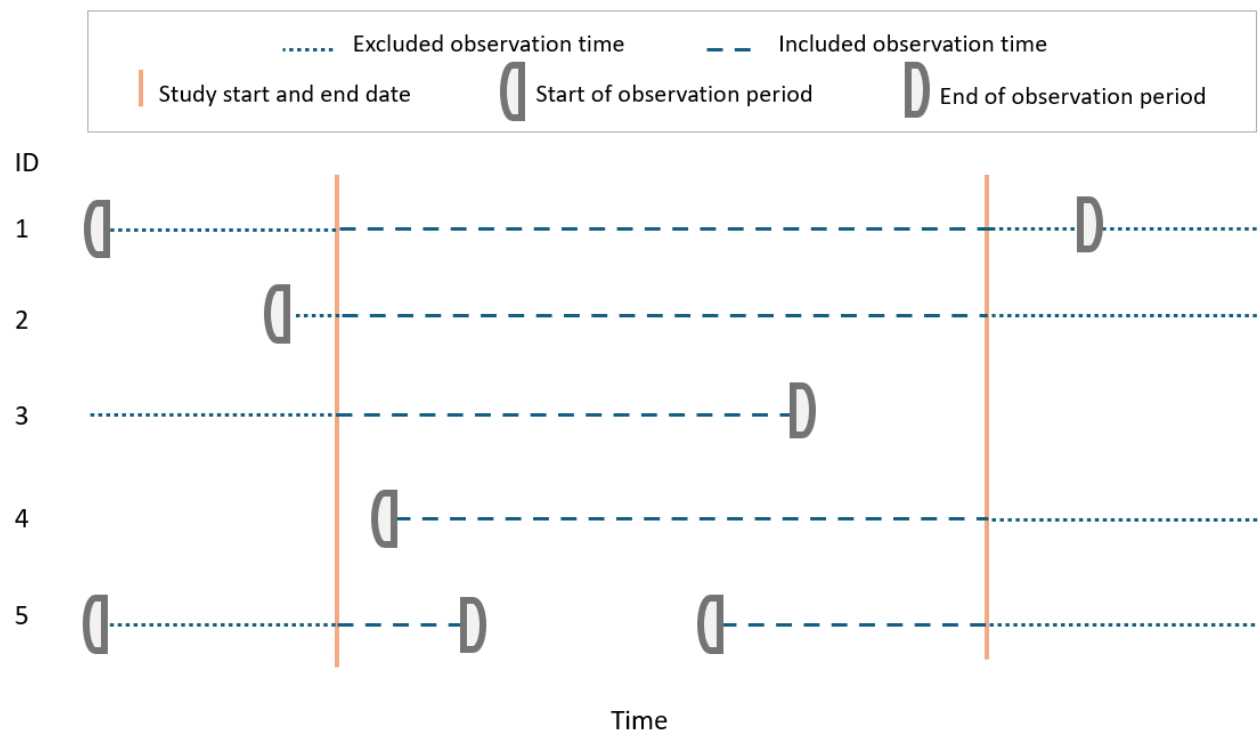


Figure S2. Included observation time for the denominator population for prevalence.

ANNEX VI. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Endometriosis and adenomyosis: Incidence, prevalence, diagnosis delay, and treatments in European countries.

EU PAS Register® number: EUPAS1000001099
Study reference number (if applicable): P5-C1-014

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

Comments:

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

Comments:

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

² Date from which the analytical dataset is completely available.

Section 3: Study design		Yes	No	N/A	Section Number
3.1	Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	8.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	8.4
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	8.8
3.4	Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☒	
3.5	Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

Comments:

Section 4: Source and study populations		Yes	No	N/A	Section Number
4.1	Is the source population described?	☒	☐	☐	8
4.2	Is the planned study population defined in terms of:				
4.2.1	Study time period	☒	☐	☐	8.5
4.2.2	Age and sex	☒	☐	☐	8.6
4.2.3	Country of origin	☒	☐	☐	8.4
4.2.4	Disease/indication	☒	☐	☐	8.6
4.2.5	Duration of follow-up	☒	☐	☐	8.2
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	8.3

Comments:

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

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

Comments:

<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?	☒	☐	☐	8.6.1
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.6.1; 8.8.3
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☐	☒	
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

<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)	☐	☐	☒	
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☐	☐	☒	
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☐	☐	☒	

Comments:

<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)	☐	☐	☒	

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)	☒	☐	☐	ANNEX I; ANNEX II
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	ANNEX I; ANNEX II
9.1.3 Covariates and other characteristics?	☒	☐	☐	ANNEX I; ANNEX II
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☒	☐	☐	ANNEX I; ANNEX II
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	ANNEX I; ANNEX II
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	ANNEX I; ANNEX II
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	ANNEX I; ANNEX II
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	ANNEX I; ANNEX II
9.3.3 Covariates and other characteristics?	☒	☐	☐	ANNEX I; ANNEX II
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☒	

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?	☒	☐	☐	8.8
10.2 Is study size and/or statistical precision estimated?	☒	☐	☐	8.7
10.3 Are descriptive analyses included?	☒	☐	☐	8.8
10.4 Are stratified analyses included?	☒	☐	☐	8.8
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	
10.7 Does the plan describe methods for handling missing data?	☐	☐	☒	
10.8 Are relevant sensitivity analyses described?	☐	☐	☒	8.8

Comments:

<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)	☒	☐	☐	ANNEX III
11.2 Are methods of quality assurance described?	☒	☐	☐	ANNEX III
11.3 Is there a system in place for independent review of study results?	☐	☐	☒	

Comments:

<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?	☒	☐	☐	
12.1.2 Information bias?	☒	☐	☐	
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☐	☐	☒	9

Section 12: Limitations	Yes	No	N/A	Section Number
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	ANNEX II

Comments:

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

Comments:

Section 14: Amendments and deviations	Yes	No	N/A	Section Number
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	4

Comments:

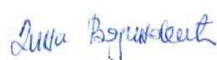
Section 15: Plans for communication of study results	Yes	No	N/A	Section Number
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	☒	☐	☐	8.8.4; ANNEX III
15.2 Are plans described for disseminating study results externally, including publication?	☒	☐	☐	8.8.4; ANNEX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

Comments:

Name of the main author of the protocol: Dina Vojinovic

Date: 03/08/2026

Signature:



ANNEX VII. 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.