Study Protocol

P4-C1-013

DARWIN EU® - Treatment patterns in postmenopausal women in European countries: A real-world observational study

23/01/2026

Version 5.0

Authors: Ellen Gerritsen, Dina Vojinovic

Public

CONTENTS

Study title	DARWIN EU® - Treatment patterns in postmenopausal women in European countries: A real-world observational study
Protocol version	V5.0
Date	23/01/2026
EUPAS number	EUPAS1000000821
Active substance	Hormone replacement therapy (HRT) including oestrogen, progestogen, and combination products.
Medicinal product	Medicinal products mapped to the pre-specified HRT. A preliminary list is provided in **ANNEX III**.
Research question and objectives	Research question: What are the patterns and trends in hormone replacement therapy (HRT) utilisation among postmenopausal women in European countries over the past three decades? Study objectives: 1. To describe the demographic and clinical characteristics of postmenopausal women, including: a) age, b) body mass index (BMI), c) smoking status, d) concomitant conditions, e) HRT used at treatment initiation, where applicable. Estimates will be stratified by HRT status and presented for selected calendar years. 2. To estimate annual prevalence of HRT use among postmenopausal women, overall and stratified by: a) HRT formulation. b) route of administration (systemic or local). c) HRT type (natural or synthetic) (if feasible). Estimates will additionally be stratified by age group at treatment initiation. 3. To estimate the duration of HRT use among postmenopausal women initiating HRT, overall, and by age group. 4. To describe HRT treatment patterns among postmenopausal women over ten years following HRT initiation, overall, and by age group.
Countries of study	Denmark, Germany, Spain, The United Kingdom
Authors	Ellen Gerritsen, e.gerritsen@darwin-eu.org Dina Vojinovic, d.vojinovic@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/term	Description
ATC	Anatomical Therapeutic Chemical
BIFAP	Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público
BMI	Body mass index
CC	Coordination centre
CDM	Common Data Model
CPRD GOLD	Clinical Practice Research Datalink GOLD
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
DOI	Declaration of Interests
DRE	Digital Research Environment
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
GDPR	General Data Protection Regulation
GP	General Practitioner
HRT	Hormone replacement therapy
ICD	International Classification of Diseases
IQVIA DA Germany	IQVIA Disease Analyzer Germany
IP	Inpatient
IQR	Interquartile range
IRB	Institutional Review Board
NA	Not applicable
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
RxNorm	Medical prescription normalized
SIDIAP	The Information System for Research in Primary Care
SNOMED	Systematized Nomenclature of Medicine
VMS	Vasomotor symptoms
WHO	World Health Organisation

1. TITLE

DARWIN EU® - Treatment patterns in postmenopausal women in European countries: A real-world observational study

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigators	Ellen Gerritsen Dina Vojinovic	IQVIA
Data Scientists	Akram Mendez Gargi Jadhav	IQVIA
Study Manager	Natasha Yefimenko	Erasmus MC
Data Partner*	**Names**	**Organisation**
DK-DHR	Elvira Bräuner Susanne Bruun	Danish Medicines Agency
IQVIA DA Germany	Isabella Kaczmarczyk James Brash	IQVIA
BIFAP	Ana Llorente-Garcia Miguel Angel Macia Martinez Marta Monreal Di Bello Cristina Justo Astorgano	Agencia Española de Medicamentos y Productos Sanitarios
SIDIAP	Anna Palomar Cros Agustina Giuliodori Picco Irene López Sánchez Laura Granés González (Invitado) Elena Roel Herranz	Institute for Primary Health Care Research Jordi Gol i Gurina
CPRD GOLD	Antonella Delmestri Marta Pineda Moncusí	University of Oxford

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

3. ABSTRACT

Title

DARWIN EU® - Treatment patterns in postmenopausal women in European countries: A real-world observational study

Rationale and background

Vasomotor symptoms (VMS), including hot flashes and night sweats, affect up to 80% of menopausal women and significantly impair quality of life. Hormone replacement therapy (HRT) remains the primary treatment for VMS, with various formulations and regimens available depending on clinical factors. Despite its widespread use, there are risks associated with HRT, which vary by age, timing, and treatment type. However, contemporary data on HRT utilisation patterns in Europe are limited, highlighting the need for real-world evidence to inform regulatory decisions.

Research question and objectives

Research question

What are the patterns and trends in hormone replacement therapy (HRT) utilisation among postmenopausal women in European countries over the past three decades?

Objectives

The specific objectives of this study are:

1. To describe the demographic and clinical characteristics of postmenopausal women, including:
    a) age,
    b) body mass index (BMI),
    c) smoking status,
    d) concomitant conditions,
    e) HRT used at treatment initiation, where applicable.

    Estimates will be stratified by HRT status and presented for selected calendar years.
2. To estimate annual prevalence of HRT use among postmenopausal women, overall and stratified by:
    a) HRT formulation.
    b) route of administration (systemic or local).
    c) HRT type (natural or synthetic) (if feasible).

    Estimates will additionally be stratified by age group at treatment initiation.
3. To estimate the duration of HRT use among postmenopausal women initiating HRT, overall, and by age group.
4. To describe HRT treatment patterns among postmenopausal women over ten years following HRT initiation, overall, and by age group.

Methods

Study design

A descriptive retrospective cohort study will be conducted using routinely collected health data from 5 data sources from 4 countries across Europe.

Population

The study population will include **postmenopausal women defined as** i) women aged 50 to 65 years and ii) women aged <50 years with a prior history of conditions or procedures indicating premature or early menopause. This definition will be applied consistently across study objectives.

Patient-level characterisation (objective 1): The study population will include postmenopausal women stratified into two cohorts: new HRT users and non-HRT users.

Population-level drug utilisation (objective 2): The study population will include all postmenopausal women who meet the criteria defined above during the study period.

Patient-level drug utilisation (objectives 3 and 4): The study population will include postmenopausal women initiating HRT during the study period.

Variables

Exposure: HRT treatment including oestrogen only, progestogen only, oestrogen + progestogen, oestrogen + testosterone, oestrogen + progestogen + testosterone, or other available formulations.

Indication: Postmenopausal symptoms/disorders

Data source

1. Denmark: Danish Data Health Registries (DK-DHR)
2. Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)
3. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
4. Spain: The Information System for Research on Primary Care (SIDIAP)
5. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Statistical analysis

Patient-level characterisation (objective 1): Clinical characteristics of interest will be reported as counts and proportions for categorical variables, and as medians with interquartile ranges (IQR) for continuous variables.

Population-level drug utilisation (objective 2): Annual prevalence of HRT use will be estimated and expressed as a proportion of women using HRT among postmenopausal women. The annual prevalence will also be estimated for HRT formulation (e.g., oestrogen only, progestogen only, oestrogen + progestogen, oestrogen + testosterone, oestrogen + progestogen + testosterone, or other available formulations), route of administration (systemic or local), and HRT type (natural or synthetic) (if feasible). All prevalence estimates will be reported overall and stratified by age group at treatment initiation (<50 years, 50–60 years, and >60 years).

Patient-level drug utilisation (objectives 3 and 4): Treatment duration will be estimated and the minimum, q25, median, q75, and maximum will be provided. The treatment patterns will be described based on the sequence of HRT treatment over ten years following HRT initiation. These analyses will be performed overall, and stratified by age group (<50 years, 50–60 years, and >60 years).

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	January 2026
Creation of Analytical code	December 2025/January 2026
Execution of Analytical Code on the data	January 2026
Draft Study Report	February 2026
Final Study Report	March 2026

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

6. RATIONALE AND BACKGROUND

Menopausal symptoms induce significant burden on the health and quality of life of women. Central among these are vasomotor symptoms (VMS; also referred to as hot flashes), typically accompanied by sweating and flushing.[1-3] Up to 80% of menopausal women experience VMS to some degree, with 20–30% reporting severe symptoms.[1-4] These vasomotor symptoms cause anxiety and sleep disruption and tend to be the most lifestyle limiting symptom in many women.[4]

Currently, there are several treatment options for oestrogen deficiency symptoms in postmenopausal women. Hormone replacement therapy (HRT) is the treatment of choice for vasomotor and genitourinary syndrome of menopause.[4] Other commonly used therapies for the relief of oestrogen deficiency symptoms are non-hormonal treatment requiring a prescription (SSRIs, SNRIs, venlafaxine, praxetine, citalopram), over the counter treatments (calcium/vitamin D, soy products), and lifestyle changes (exercise, diet).[2, 4] Treatment options are often combined to relieve symptoms.[2, 4]

HRT may contain oestrogen alone or an oestrogen and a progestogen combination which can be taken continuously or sequentially for 14 days. The choice of combining therapy with a progestogen is dependent on the presence of an intact uterus.[5-7] Oestrogens in HRT are available as oral conjugated equine oestrogens (CEE), synthetic ethinyl oestradiol, micronized oestradiol, and oestradiol valerate.[6] It is worth noting that the availability of the different forms varies between countries. In some countries the most widely prescribed oestrogen formulations might be oral preparations and, in others, transdermal oestradiol.[8] In some cases, testosterone is added to oestrogen and progestogen therapy to address symptoms, such as reduced libido and sexual dysfunction, in postmenopausal women.[9]

Some studies have suggested that the absolute risks of adverse events in postmenopausal women receiving HRT depend on several factors, including age at initiation and timing in relation to menopause, HRT type, dosage, duration, and route of administration.[4, 10-13]

To our knowledge, studies describing contemporary patterns and trends of HRT utilisation in European countries are very scarce. A better understanding of the use of these medicines in real-world settings can help inform regulatory decision making.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

What are the patterns and trends in hormone replacement therapy (HRT) utilisation among postmenopausal women in European countries over the past three decades?

Research objectives

The specific objectives of this study are:

1. To describe the demographic and clinical characteristics of postmenopausal women, including:
    a) age,
    b) body mass index (BMI),
    c) smoking status,
    d) concomitant conditions,
    e) HRT used at treatment initiation, where applicable.

    Estimates will be stratified by HRT status and presented for selected calendar years.
2. To estimate annual prevalence of HRT use among postmenopausal women, overall and stratified by:
    a) HRT formulation.
    b) route of administration (systemic or local).
    c) HRT type (natural or synthetic) (if feasible).

    Estimates will additionally be stratified by age group at treatment initiation.
3. To estimate the duration of HRT use among postmenopausal women initiating HRT, overall, and by age group.
4. To describe HRT treatment patterns among postmenopausal women over ten years following HRT initiation, overall, and by age group.

8. RESEARCH METHODS

8.1. Study design

A retrospective cohort study will be conducted using routinely collected health data from 5 data sources from 4 European countries belonging to 3 EU member states. The study will comprise of:

- A patient-level characterisation will be conducted to address *objective 1*, describing the demographic and clinical characteristics of postmenopausal women, stratified into two cohorts: new HRT users and non-HRT users (**Figure 1a**). For new HRT users, the type of HRT used at treatment initiation will also be described.
- A population-level drug utilisation study will be conducted to address *objective 2*, to estimate the annual prevalence of HRT use among postmenopausal women, overall and stratified by formulations, route of administration, and type of HRT (if feasible) (**Figure 1b**).
- A patient-level drug utilisation study will be conducted to address *objectives 3 and 4*, estimating HRT duration among postmenopausal women, and describing treatment patterns (**Figure 1c**).

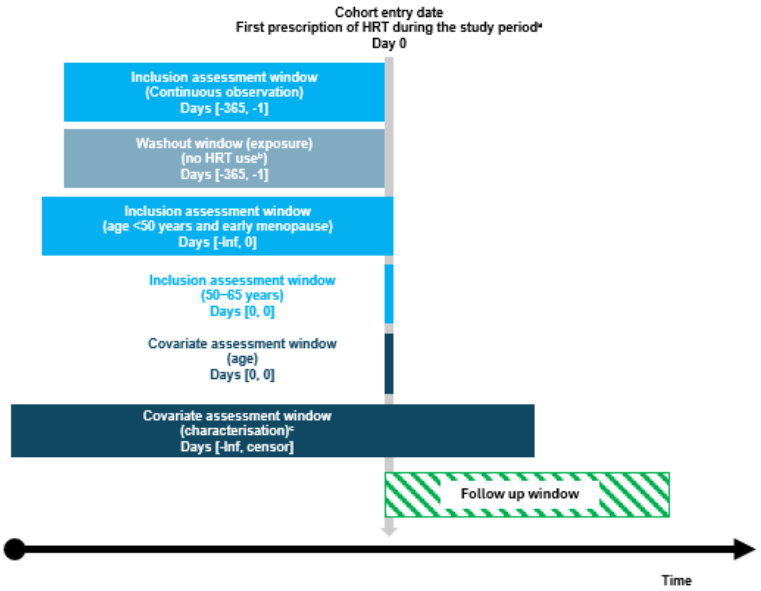

Figure 1a. Graphical depiction of the study design for objective 1.

a. Characterisation will be done for non-HRT users as well. Non-HRT users: women aged 50 years or older with no record of HRT prescription prior to the cohort entry date [-Inf, 0]. These women will be followed up until the earliest of HRT prescription record, loss to follow-up, death, end of observation period (the latest available data). They may contribute to both cohorts (HRT user and non-users). Their person time will be attributed accordingly.
Characterisation for HRT users and non-HRT users will be reported for selected calendar years.
b. Applies to women aged 50 to 65 years, and to women aged <50 years with a prior history of conditions or procedures indicating premature or early menopause initiating HRT.
c. Clinical characteristics will be described including body mass index (BMI), smoking status, menopause, and concomitant conditions. For new- HRT users, the type of HRT used at treatment initiation will also be described.

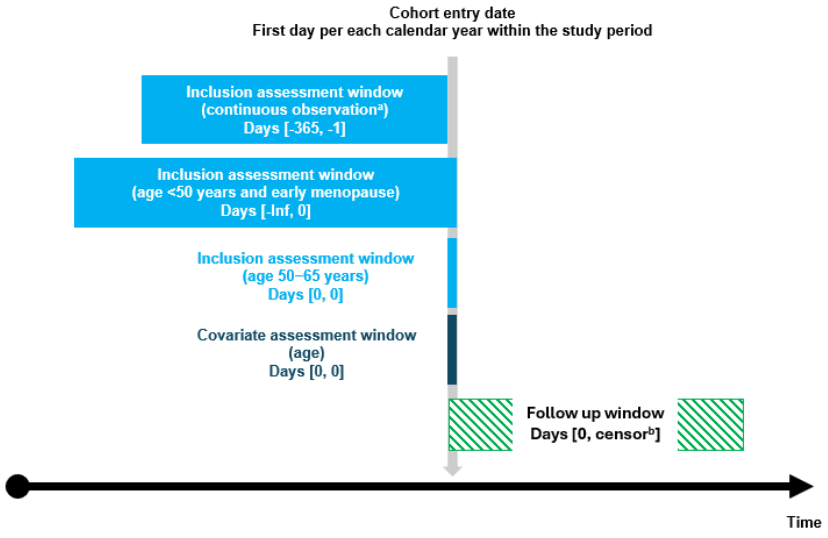

Figure 1b. Graphical depiction of the study design for objective 2.

a. Applies to women aged 50 to 65 years, and to women aged <50 years with a prior history of conditions or procedures indicating premature or early menopause.
b. Death, loss to follow-up, end of data source availability, end of each calendar year (i.e., 31 December), or end of the study period (31/12/2024).

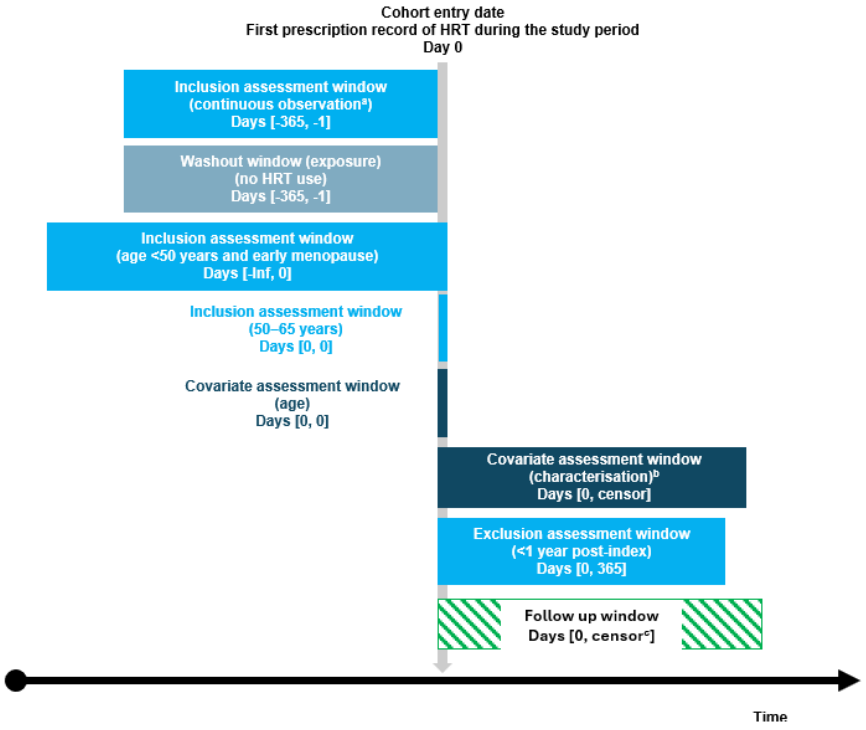

Figure 1c. Graphical depiction of the study design for objectives 3 and 4.

a. Applies to women aged 50 to 65 years, and women aged <50 years with a prior history of conditions or procedures indicating premature or early menopause.
b. Characterisation of HRT treatment pattern.
c. Death, loss to follow-up, end of data source availability, or end of the study period (31/12/2024).

8.2. Study setting and data sources

This study will be conducted using routinely collected data from 1 registry and 4 primary care data sources of the DARWIN EU® data network from 4 European countries covering 3 EU member states. All data have been a priori mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

Data sources

1. Denmark: Danish Data Health Registries (DK-DHR)
2. Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)
3. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
4. Spain: The Information System for Research on Primary Care (SIDIAP)
5. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Data Selection

These data sources fulfil the criteria required in terms of data quality, completeness, timeliness, and representativeness for drug utilisation study while covering different regions of Europe (**ANNEX I**).

Data source justification and key characteristics

Multiple data sources will be included in this study, as they represent registry-based and outpatient general practitioner care databases that provide relevant information on HRT use in postmenopausal women. Based on preliminary feasibility assessments, the expected number of individuals with oestrogen use

ranges from approximately 2,100 (SIDIAP) to 251,800 (CPRD GOLD), while the number of individuals with oestrogen + progesterone product use ranges from 800 (CPRD GOLD) to 254,800 (DK-DHR). Data availability and follow-up are sufficient to support the study objectives, with data collection start years ranging from 1987 (CPRD GOLD) to 2006 (SIDIAP), and the most recent data releases occurring between November 2024 (DK-DHR) and July 2025 (CPRD GOLD). Median follow-up duration varies across sources, ranging from 116 days (IQVIA DA Germany) to 7,920 days (DK-DHR), ensuring adequate longitudinal coverage. While IQVIA DA Germany and CPRD GOLD align with the intended study period (1 January 1994 to 31 December 2024), for these data sources we will use the earliest date available. The actual study period for other data sources is adjusted based on data quality considerations, specifically regarding the earliest reliable data availability and the last observation period available. Subsequently, the actual study period will be from 1995 until November 2024 in DK-DHR (the last date of prescription data is 30 September 2024), from 2005 until December 2024 in BIFAP, and from 2010 until December 2024 in SIDIAP. Institutional Review Board (IRB) approvals are either covered under existing umbrella protocols or expected to be obtained within 1 to 3 months (BIFAP, SIDIAP), supporting timely execution of the study within the established timelines.

Information on the data sources planned for use in this study is provided in **Table 1** and **ANNEX I.**.

Table 1. Data source justification and key characteristics.

Country	Data source acronym	Type of data	Oestrogens (n)[a]	Oestrogens + progesterone (n)[a]	Data collection start year	Source release date	Last observation period month	Follow-up duration (days), median (IQR)[b]	IRB approval
Denmark	DK-DHR	Registry	2,700	254,800	1995	18/01/2025	11/2024	7,920 (2,610–10,900)	Umbrella protocol
Germany	IQVIA DA Germany	Outpatient General Practitioner Care	29,400	36,500	1992	31/12/2024	06/2025	116 (0–1,610)	Not needed
Spain	BIFAP	GP and hospital care	3,800	-	2001 (good quality 2005)	31/12/2024	12/2024	2,550 (2,050–5,430)	1–3 months
Spain	SIDIAP	Outpatient General Practitioner Care	2,100	33,000	2006 (good quality from 2010)	30/06/2023	12/2024	5,670 (2,220–6,390)	1–3 months
UK	CPRD GOLD	Outpatient General Practitioner Care	251,800	800	1987	01/01/2025	07/2025	2,150 (727–4,930)	Umbrella protocol

Abbreviations: DK-DHR = Danish Data Health Registries; IQVIA DA Germany = IQVIA Disease Analyzer Germany; BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público; SIDIAP = The Information System for Research on Primary Care; CPRD GOLD = Clinical Practice Research Datalink GOLD; GP = general practitioner; IQR = interquartile range.
[a] = person counts at drug era (as generated from drug era).
[b] = median follow-up of the first observation period.

8.3. Study period

The study period is from 1 January 1994 (or the earliest date available with data of sufficient quality) to 31 December 2024 (or the most recent data available) for each contributing data source.

It should be noted for several data sources, the availability of the accurate data deviates from the start or end date of the study period. Detailed information about the study period per data partner can be found in **Section 8.2**.

8.4. Follow-up

For characterisation of postmenopausal women initiating HRT treatment *(objective 1)*, patient-level drug utilisation of treatment duration *(objective 3)* and description of treatment patterns *(objective 4),* follow-up will start on the date of the first recorded prescription of HRT during the study period, provided that the woman meets the postmenopausal definition (either by age 50–65 or <50 with a history of premature/early menopause), she has no prior HRT exposure in 365 days before this date and has at least 365 days of data visibility prior to this date. End of follow-up will be defined as the earliest of loss to follow-up, death, or end of observation period (the latest available data) for *objectives 1* and *3.* For *objective 4,* follow-up will additionally be censored at 10 years after the date of HRT initiation.

For characterisation of postmenopausal women not on HRT *(objective 1)*, follow-up will start once women meet the inclusion criteria (age 50 years or older with no record of prior HRT prescription at any time prior to the cohort entry and at least 365 days of data visibility). End of follow-up will be defined as the earliest of HRT prescription record, loss to follow-up, death, end of observation period (the latest available data). Women may contribute to both cohorts (HRT users and non-users). Their person time will be attributed accordingly.

For calculation of population-level drug utilisation *(objective 2),* follow-up will start on the respective date of the latest of the following: i) study start date 1 January 1994 (or earliest date available with data of sufficient quality), ii) date at which they have 365 days of prior history, provided that women meet the postmenopausal definition (either by age 50–65 or <50 with a history of premature/early menopause). End of follow-up will be defined as the earliest of loss to follow-up, death, or end of observation period (the latest available data), whatever occurs first.

An example of entry and exit into the denominator population is shown in **Figure 2**. 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 person 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.

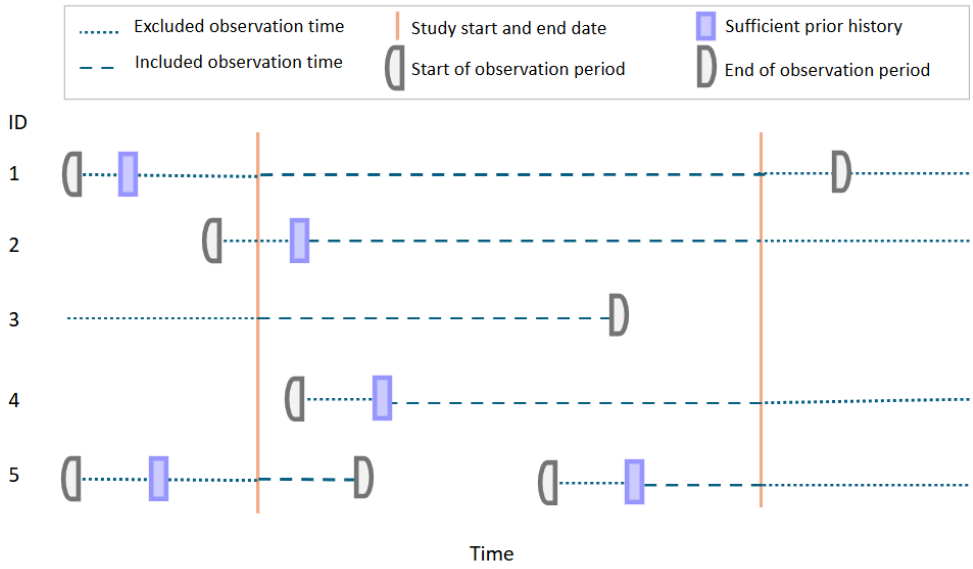

Figure 2. Included observation time for the denominator population.

8.5. Study population with inclusion and exclusion criteria

For characterisation of postmenopausal women initiating HRT treatment *(objective 1)*, patient-level drug utilisation to estimate treatment duration *(objective 3)*, and description of treatment patterns *(objective 4),* the study population will include postmenopausal women initiating HRT during the study period.

Inclusion criteria:

- Postmenopausal women defined as i) women aged 50 to 65 years and ii) women aged <50 years with a prior history of conditions or procedures indicating premature or early menopause (primary ovarian insufficiency/failure, premature menopause, early menopause, bilateral oophorectomy, or hysterectomy with ovarian removal).
- First record of HRT prescription between 1 January 1994 (or the earliest available with data of sufficient quality) to 31 December 2024 (or the most recent data available).
- At least 365 days of data visibility prior to HRT initiation.
- No prior HRT exposure in 365 days prior to HRT initiation.

Exclusion criteria:

- Postmenopausal women initiating HRT <365 days before the end of data availability in the respective data source (to ensure sufficient follow-up) *(objectives 3* and *4).*

The preliminary concept sets used for the identification of premature or early menopause are described in **Table S2** of **ANNEX III**.

For characterisation of postmenopausal women not on HRT *(objective 1),* the study population will include:

Inclusion criteria:

- Women aged 50 years or older.
- No record of HRT prescription prior to or at the cohort entry date [-Inf, 0].
- At least 365 days of data visibility prior to the cohort entry date.

Exclusion criteria:

No formal exclusion criteria are applied beyond the inclusion criteria.

For calculation of prevalence of HRT use *(objective 2)*, the study population will include postmenopausal women.

Inclusion criteria:

- Postmenopausal women defined as i) women aged 50 to 65 years and ii) women aged <50 years with a prior history of conditions or procedures indicating premature or early menopause (primary ovarian insufficiency/failure, premature menopause, early menopause, bilateral oophorectomy, or hysterectomy with ovarian removal).
- At least 365 days of data visibility prior to the index date.

Exclusion criteria:

No formal exclusion criteria are applied beyond the inclusion criteria.

8.6. Variables

8.6.1. Exposure

The primary exposure is defined as a prescription or dispensing record of HRT, representing the first recorded prescription in the study period with no prior HRT use in the 365 days prior to the index date.

To define treatment episodes, sequential prescriptions will be grouped into drug eras, allowing a maximum gap of 90 days between the end of one prescription and the start of the next.

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

8.6.2. Outcome

The outcomes are as follows:

- Characterisation of postmenopausal women initiating HRT treatment or not on HRT *(objective 1).*
- Annual prevalence of HRT use among postmenopausal women, including formulations, route of administration, and type of HRT (if feasible) (*objective 2).*
- Duration of HRT treatment among postmenopausal women *(objective 3).*
- Treatment patterns following HRT treatment initiation *(objective 4).*

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

The variables for characterisation of postmenopausal women initiating HRT treatment or not on HRT treatment *(objective 1)* are as follows:

- Demographic and clinical characteristics:
    - Age groups (<50 years, 50–60 years, and >60 years)
    - BMI (reported as median with interquartile ranges)
    - Smoking status (reported as smoker/non-smoker/missing if feasible)

  - Received menopause diagnosis (yes/no)
  - Type of HRT used at treatment initiation (where applicable) using the same HRT variables defined for treatment pattern analysis in *objective 4.*

- Concomitant conditions (reported as N (%)):
  - Cardiovascular disease
  - Hypertension
  - Diabetes mellitus
  - Dyslipidaemia
  - Metabolic syndrome
  - Osteoporosis
  - Depression
  - Anxiety
  - Sleep disorders
  - Uterus status:
    - Hysterectomised
    - Intact uterus*

*In real-world data, an intact uterus is rarely explicitly recorded. Unless a hysterectomy procedure or a diagnosis code indicating absence of uterus is present in the patient's history, the uterus is assumed to be intact.

Characterisation will be performed on 1 July for the following calendar years: 1990, 1995, 2000, 2005, 2010, 2015, 2020, and 2024.

Assessment windows:

- Demographic characteristics (age): on calendar date [0, 0]
- Clinical characteristics: 1 year prior to the calendar date [-365, 0]
- Concomitant conditions: any time prior to the calendar date [-Inf, 0] and one year post the calendar date [1, 365]

Large scale characterisation – the top 10 most frequent diagnostic codes (comorbidities) or prescribed medications (comedication) will be assessed at the pre-selected mid-year calendar date [0,0], 1 year prior to this calendar date [-365, 0] and 1 year post this date [1, 365].

The covariates considered in the analyses estimating prevalence of HRT *(objective 2)* are as follows:

- Calendar year
- Age groups: <50 years, 50–60 years, and >60 years
- HRT formulation: oestrogen only, progestogen only, oestrogen + progestogen, oestrogen + testosterone, oestrogen + progestogen + testosterone, or other available formulations
- Route of administration: systemic (oral tablets, transdermal patches, subcutaneous implants, topical sprays, gels, and creams), local (vaginal creams, tablets, rings)
- HRT type: natural, synthetic (if feasible)

The covariates for calculation of treatment duration of HRT *(objective 3)* will include:

- Age groups (<50 years, 50–60 years and >60 years)

The variables for characterisation of treatment patterns *(objective 4)* will include current and historical HRT, capturing changes over the past three decades. These variables include:

- Treatment:
    - G03CA Natural and semisynthetic oestrogens
        - Oestradiol
        - Oestriol
        - Conjugated equine oestrogens
        - Esterified oestrogens
        - Chlorotrianisene
        - Ethinylestradiol
        - Oestrone
        - Promestriene
    - G03CB Synthetic oestrogens
        - Dienestrol
        - Methallenestril
        - Moxestrol
        - Diethylstilbestrol
    - G03CC Oestrogen combinations
        - Conjugated oestrogens and bazedoxifene
    - G03CX Other oestrogens
        - Tibolone
    - G03D Progesterones
        - Micronised progesterone
        - Hydroxyprogesterone
        - Dydrogesterone
        - Norethisterone
        - Levonorgestrel (IUD)
        - Medroxyprogesterone acetate
        - Lynestrenol
        - Megestrol/Medrogestone
        - Nomegestrol
    - G03EA Combination oestrogen and androgen

        - Oestradiol; Testosterone
        - Esterified oestrogens; methyltestosterone
        - Conjugated oestrogens; methyltestosterone
        - Testosterone; oestrogen
        - Prasterone; oestrogen
    - G03F Progesterone and oestrogens, fixed combinations
        - Oestradiol; Levonorgestrel
        - Oestradiol; Medroxyprogesterone
        - Oestradiol; Norethisterone
        - Oestradiol; Dydrogesterone
        - Oestradiol; Progesterone
        - Oestradiol; Drospirenone
        - Conjugated oestrogens; Medroxyprogesterone
    - Combination androgen, progesterone, and oestrogens
        - Oestradiol; Norethindone; Testosterone
- Stratification:
    - Age groups: <50 years, 50–60 years and >60 years

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

8.7. Study size

No sample size has been calculated as this is a drug utilisation study with a descriptive design and no formal hypothesis testing. The study will rely on existing routinely collected data to estimate the prevalence of HRT use and to describe treatment patterns. Thus, the sample size is driven by the availability of data for eligible individuals within the included data sources.

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 patient-level data.

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

The data partners will locally execute the analytics against the OMOP CDM in R Studio and review and approve the default aggregated results. They will then be made available to the Principal Investigators and study team in a secure online repository (Data Transfer Zone). The study results of all data sources are

checked, after which they are made available to the team, and the Study Dissemination Phase can start. All results are locked and timestamped for reproducibility and transparency.

8.8.2. Patient privacy protection

All analyses will be conducted separately for each data source and will be carried out in a federated manner, allowing analyses to be run locally without sharing patient-level data. Cell counts <5 will be suppressed when reporting results to comply with the data source's privacy protection regulations.

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

R-packages

The patient-level characterisation of postmenopausal women, initiating HRT treatment or not on HRT *(objective 1)*, will be estimated based on OMOP CDM mapped data using the *CohortCharacteristics* R package (https://darwin-eu.github.io/CohortCharacteristics/). Population-level utilisation of HRT in postmenopausal women *(objective 2)* will be assessed using the *IncidencePrevalence* R package (https://darwin-eu.github.io/IncidencePrevalence/), the patient-level utilisation including treatment duration *(objective 3)* will be estimated using the *DrugUtilisation* R package (https://github.com/darwin-eu/DrugUtilisation), and patient-level characterisation of treatment patterns *(objective 4)* will be described using the *TreatmentPatterns* R package (https://github.com/darwin-eu-dev/TreatmentPatterns). All packages are publicly available and developed by DARWIN EU®.

Characterisation of postmenopausal women *(objective 1)*

Postmenopausal women, including those initiating HRT or those not on HRT, will be characterised in terms of age, clinical characteristics (BMI, smoking status, code for menopause), and concomitant conditions. Clinical characteristics will be reported as counts and proportions for categorcial variables and as medians with interquartile ranges (IQR) for continuous variables. For clinical characteristics, the percentage of unknowns will also be reported. Among women initiating HRT, the type of treatment used at initiation will also be described using counts and proportions.

Prevalence of HRT use in postmenopausal women *(objective 2)*

Prevalence will be calculated as annual period prevalence, which summarises the total number of postmenopausal women who are using HRT (*objective 2*) during a given calendar year divided by the total number of postmenopausal women at risk of initiating HRT in that year. This measure reflects the proportion of individuals exposed to HRT during a specified time interval. Binomial 95% confidence intervals will be computed to quantify statistical uncertainty around the prevalence estimates.

An illustration of the calculation of period prevalence is shown below in **Figure 3**.**Error! Reference source not found.** Between time t+2 and t+3, two of the five study participants are users of the pre-selected drug of interest, giving a prevalence of 40%. Meanwhile, for the period t to t+1, all five also have some observation time during the year, with one of the five study participants being a user of the pre-selected drug of interest.

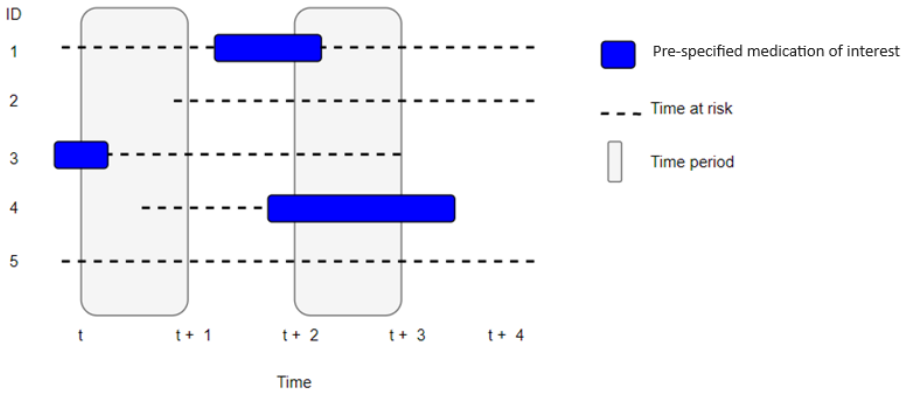

Figure 3. Period prevalence example.

Annual prevalence estimates will be presented overall, and by age group (<50 years, 50–60 years, >60 years). If applicable, age at the index date will be calculated using January 1 of the year of birth as proxy for the actual date of birth. Date/month is either not present or cannot be made available for governance reasons. If available, date is often set to first of the month for patient's privacy.

Drug exposure calculations

Drug eras will be defined as follows: exposure to HRT starts at the date of the new prescription within the study period after a washout interval of 365 days. For each prescription, the estimated duration of use is retrieved from the drug exposure table in the CDM, utilising both the start and end dates of the exposure. Subsequent prescriptions will be combined into continuous exposed episodes (drug eras) using the following specifications: two drug eras will be merged into one continuous drug era if the distance in days between the end of the first era and the start of the second era is ≤ 90 days. To define treatment episodes, we apply the gap era, which specifies the maximum allowable gap (in days) between consecutive treatment records for them to be considered part of the same episode. As illustrated in **Figure 4**, setting the gap era to 0 will result in three distinct episodes, joining only overlapping records. When the gap era is set to any value between 1 and 28 days, consecutive episodes separated by gaps within this range will be merged, resulting in two episodes. If the gap era exceeds 29 days, all records will be collapsed into a single continuous episode. For the purpose of this study, the gap era will be set to ≤ 90 days. Consequently, prescription records corresponding to drug exposures 1, 2, 3, and 4 will be aggregated into a single continuous episode.

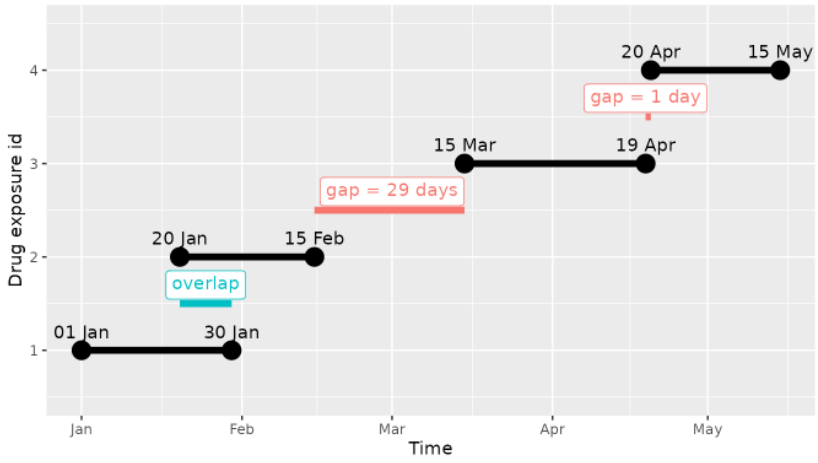

Figure 4. Gap era joint mode.

Bold horizontal black lines represent drug episodes recorded at different times. Prescribing/dispensing records of the same drug are illustrated on the Y-axis and time in months on the X-axis.

If two exposures overlap, the overlap time will be considered exposed to the first exposure. No time will be added at the end of the combined drug era to account for the overlap. If two exposures start at the same date, the overlapping period will be considered exposed to both.

Treatment duration *(objective 3)*

Treatment duration will be calculated as the total number of days an individual is continuously exposed to HRT among postmenopausal women. Exposure periods will be constructed by merging consecutive prescription records, allowing for treatment gaps of up to 90 days between prescriptions. The duration will be calculated from the start date of the first prescription to the end date of the last prescription within the merged sequence.

In accordance with OMOP CDM conventions, treatment duration is derived from drug_exposure_start_date and drug_exposure_end_date, with an additional day included to account for both the start and end dates in the exposure period (https://ohdsi.github.io/CommonDataModel/cdm54.html#drug_exposure/). This calculation assumes accurate recording of both dates, as discrepancies could affect the estimated duration of exposure time. Continuous exposure is assumed throughout the interval between the start and end dates, allowing for a maximum interruption of 90 days. Periods of drug exposure before the cohort start or after the cohort end date will be excluded. Missing exposure data will be treated as zero, indicating no exposure during periods with missing information.

In cases where the end date is not explicitly available in the data, it may be inferred during the ETL process using one of the following methods: (1) adding the value of treatment duration or days' supply to the start date, (2) using quantity divided by daily dose or concentration, or (3) applying predefined rules such as assigning a default value of 29 days (which is a predefined OMOP convention) for written prescriptions or +89 days for mail-order prescriptions.

Treatment duration will be summarised using descriptive statistics, including minimum, quartiles, and maximum values. Where duration cannot be reliably calculated, it will not be reported.

Treatment patterns (*objective 4*)

The number and percentage of patients receiving each of the treatment options (**Section 8.6.3.**), as well as treatment combinations, will be described in the study population (*objective 4*). Treatment

combinations will be defined algorithmically based on the drug input cohorts (**Section 8.6.3.**) and pre-specified parameters (**Figure 5**).

Additionally, Sunburst plots and Sankey diagrams will be used to visualise treatment patterns and sequences over time. Sankey diagrams will be censored as described in **Section 8.4**.

To construct treatment pathways, various parameters can be defined in the *TreatmentPatterns* package (**Figure 5**). A summary of decisions to construct individual treatment pathways is provided below. In this context, events labelled A, B, and C refer to the treatments of interest. For a comprehensive description of the steps involved, from extracting information from medical records to generating treatment pathways, please refer to the full methodology in the referenced publication.[14]

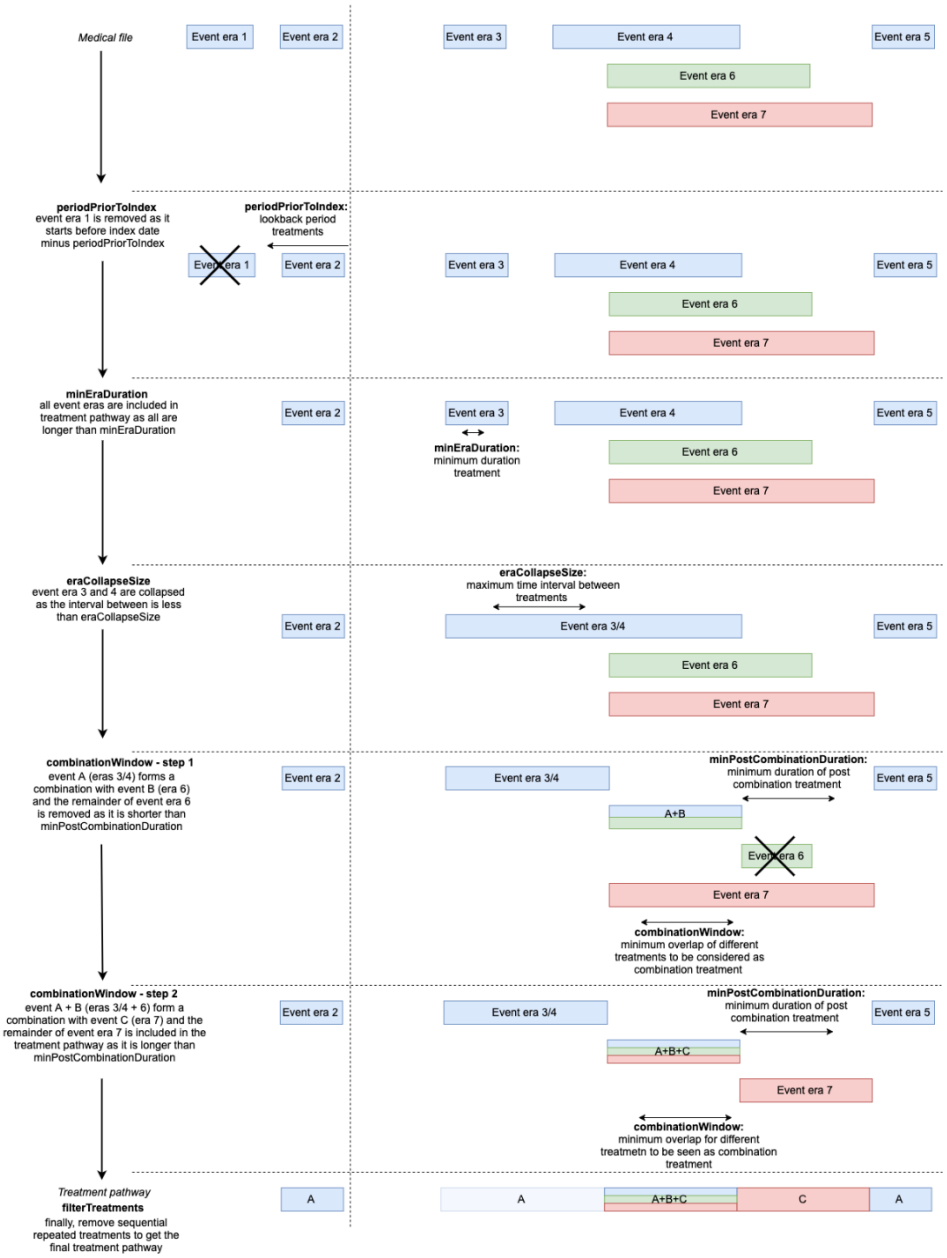

Figure 5. Parameters in the *TreatmentPatterns* package.

The preliminary parameters outlined in this study are described in **Table 2**. Preliminary treatment pattern parameters are defined based on existing knowledge and will be refined following drug exposure

diagnostics assessing episode length distributions. HRT is a chronic therapy typically prescribed in monthly or multi-month cycles; therefore, an era collapse size of 90 days accommodates realistic refill gaps without misclassifying temporary interruptions.

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

Individual pathway settings		
periodPriorToIndex	The period (number of days) prior to the index date of the target cohort from which treatments should be included	0
minEraDuration	Minimum time an event era should last to be included in the analysis	30
eraCollapseSize	Maximum gap within two eras of the same event cohort which would still allow the eras to be collapsed into one era	90
combinationWindow	Time that two event eras need to overlap to be considered a combination treatment	30
minPostCombinationDuration	Minimum time that an event era before or after a generated combination treatment should last to be included in the pathway as a separate treatment	30
filterTreatments	Select which treatments should be included in pathway: first time occurrences of treatments ('First'), remove sequential repeated treatments ('Changes'), all treatments ('All')	Changes
maxPathLength	Maximum number of treatments included in pathway	5
Aggregate pathway settings (post-processing)*		
minCellCount	Minimum number of persons with a specific treatment pathway for the pathway to be included in analysis	5
minCellMethod	Select to completely remove / sequentially adjust (by removing last step as often as necessary) treatment pathways below minCellCount	Adjust
groupCombinations	Select to group all non-fixed combinations in one category 'other' in the sunburst plot	TRUE
addNoPaths	Select to include untreated persons without treatment pathway in the sunburst plot	TRUE

*Aggregate pathway settings are post-processing parameters applied after individual patient-level pathways have been generated. They do not affect the underlying treatment episodes or sequence construction. Instead, they control how pathways are summarised, filtered, and visualised in aggregate outputs.

Method to deal with missing data

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

Sensitivity analysis

Sensitivity analyses are presented by means of **Table 3**. To assess the robustness of prevalence findings (*objective 2)* and treatment patterns (*objective 4)*, study objectives will be reevaluated using a narrower population definition, which includes: 1) women aged 50 to 65 years with a diagnosis of menopause or menopause-specific symptoms and 2) women under 50 years with a history of conditions or procedures indicating premature or early menopause.

Table 3. Sensitivity analyses – rationale, strengths, and limitations.

Sensitivity analysis	What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Assessment of narrow study population definition	The definition of the study population varies. The sensitivity population adds a clinical confirmation layer by requiring a diagnosis of menopause or menopause-specific symptoms for women aged 50–65.	To test the robustness of findings.	Greater specificity in identifying true postmenopausal women.	Women with documented menopause diagnoses or symptoms may differ from others in healthcare use or comorbidities, which could introduce bias. This analysis relies heavily on the accuracy and completeness of diagnostic coding. If menopause-related conditions are under-recorded or inconsistently coded across data sources, the sensitivity population may not fully represent the intended subgroup.

8.8.4. Output

Output will include a PDF report including an executive summary and the following mock tables and figures:

Table 1. Mock table shell - Attrition table.

Variable	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5
Qualifying initial record[1]					
Cohort start date between 1/1/1994 (or earliest available) and 31/12/2024 (or latest available)					
Record of HRT prescription					
365 days of prior data visibility					
No prior HRT exposure in 365 days					

[1]Refers to postmenopausal women defined as i) women aged 50 to 65 years, and ii) women aged <50 years with a prior history of conditions or procedures indicating premature or early menopause.

Table 2. Mock table shell - Age of postmenopausal women stratified by HRT use across selected calendar years by data source *(objective 1)*.

	N	HRT initiators			non-HRT users		
		<50 y	50–60	>60	<50 y	50–60	>60
Data source 1							
1990							
1995							
2000							
2005							
2010							
2015							
2020							
2024							
Data source 2							
1990							

Table 3. Mock table shell - Clinical characteristics of postmenopausal women stratified by HRT use across selected calendar years by data source *(objective 1)*.

Variable	N	1990	1995	2000	2005	2010	2015	2020	2024
Data source 1									
	HRT initiators								
Clinical characteristics									
• BMI (kg/m^2)									
• Smoking									
• Had menopause diagnosis									

Variable	N	1990	1995	2000	2005	2010	2015	2020	2024
Concomitant conditions									
• Cardiovascular disease									
• Hypertension									
• Diabetes mellitus									
• Dyslipidaemia									
• Metabolic syndrome									
• Osteoporosis									
• Depression									
• Anxiety									
• Sleep disorders									
Uterus status:									
• Hysterectomised									
• Intact uterus									
HRT treatment at initiation									
• Oestradiol									
• Oestriol									
• Other HRT treatment types (full list in Section 8.6.3.)									
Non-HRT users									
Clinical characteristics									
• BMI (kg/m^2)									
• Smoking									
• Had menopause diagnosis									
Concomitant conditions									

Variable	N	1990	1995	2000	2005	2010	2015	2020	2024
• Cardiovascular disease									
• Hypertension									
• Diabetes mellitus									
• Dyslipidaemia									
• Metabolic syndrome									
• Osteoporosis									
• Depression									
• Anxiety									
• Sleep disorders									
Uterus status:									
• Hysterectomised									
• Intact uterus									
HRT treatment at initiation									
• Oestradiol									
• Oestriol									
• Other HRT treatment types (full list in Section 8.6.3.)									

Table 4. Mock table shell - Duration of HRT treatment among postmenopausal women, overall and stratified by age group *(objective 3).*

	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5
Overall					
Duration (median, IQR)					
Duration (min, max)					
Age group <50 years					
Duration (median, IQR)					
Duration (min, max)					
Age group 50–60 years					
Duration (median, IQR)					
Duration (min, max)					
Age group >60 years					
Duration (median, IQR)					
Duration (min, max)					

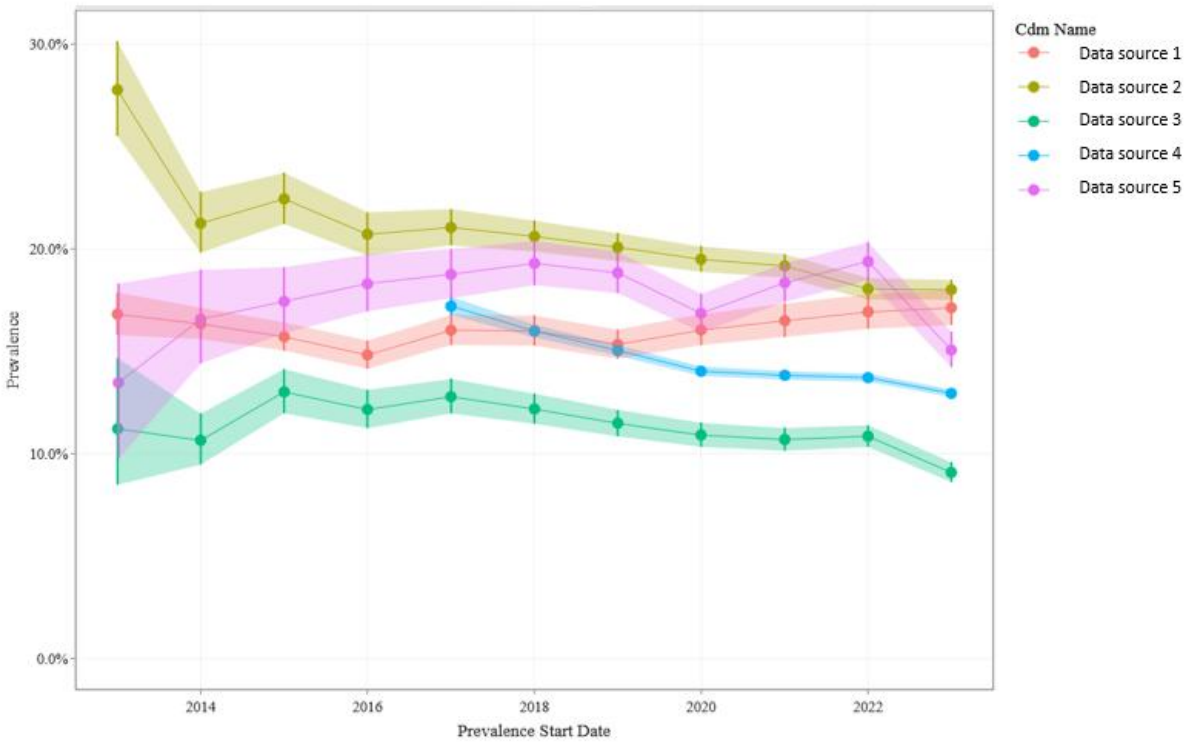

Figure 1. Mock figure - Annual prevalence of HRT use among postmenopausal women *(objective 2)*.

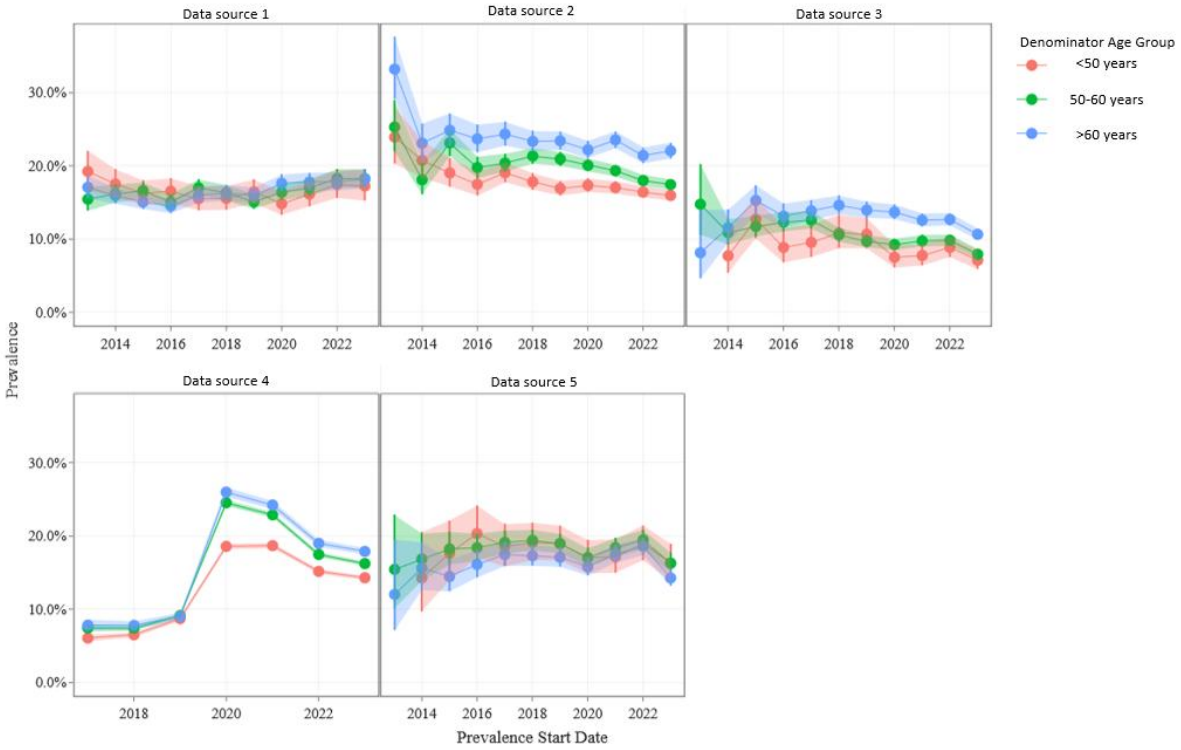

Figure 2. Mock figure - Annual prevalence of HRT use among postmenopausal women, stratified by age group *(objective 2)*.

Figures 1 and 2 are provided as mock visualisations to illustrate the intended graphical format. The figures listed below (Figures 3–8) are not shown but will follow a similar structure and visual style, tailored to each objective and data output.

Figure 3. Mock figure – Annual prevalence by HRT formulation *(objective 2)*.

Figure 4. Mock figure – Annual prevalence by HRT formulation, stratified by age group *(objective 2)*.

Figure 5. Mock figure – Annual prevalence by route of administration *(objective 2)*.

Figure 6. Mock figure – Annual prevalence by route of administration, stratified by age group *(objective 2)*.

Figure 7. Mock figure – Annual prevalence by HRT type *(objective 2)*.

Figure 8. Mock figure – Annual prevalence by HRT type, stratified by age group *(objective 2)*.

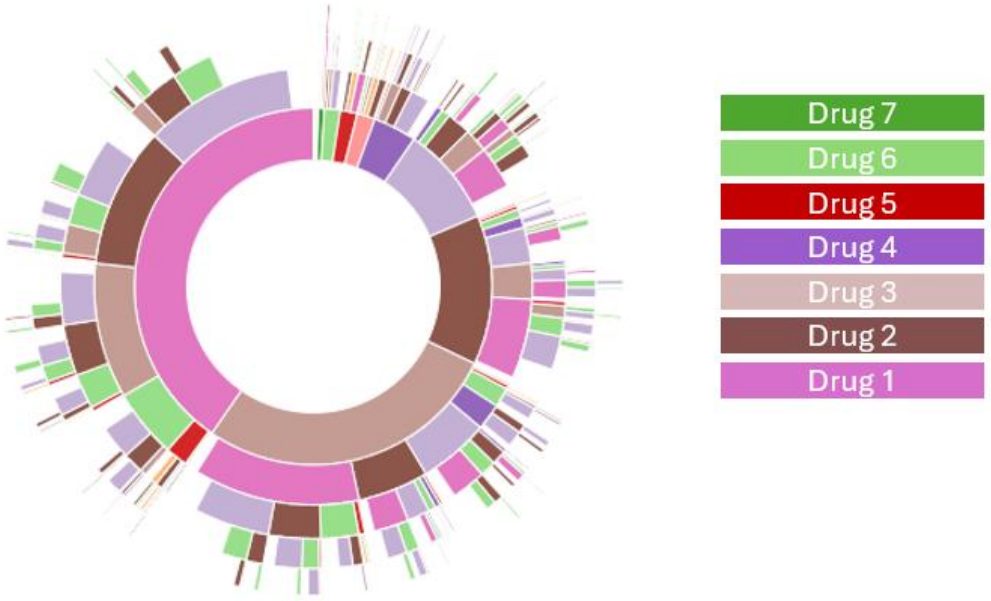

Figure 9. Sunburst plot depicting treatment patterns over 10 years following HRT initiation in postmenopausal women, overall in data source 1 *(objective 4).*

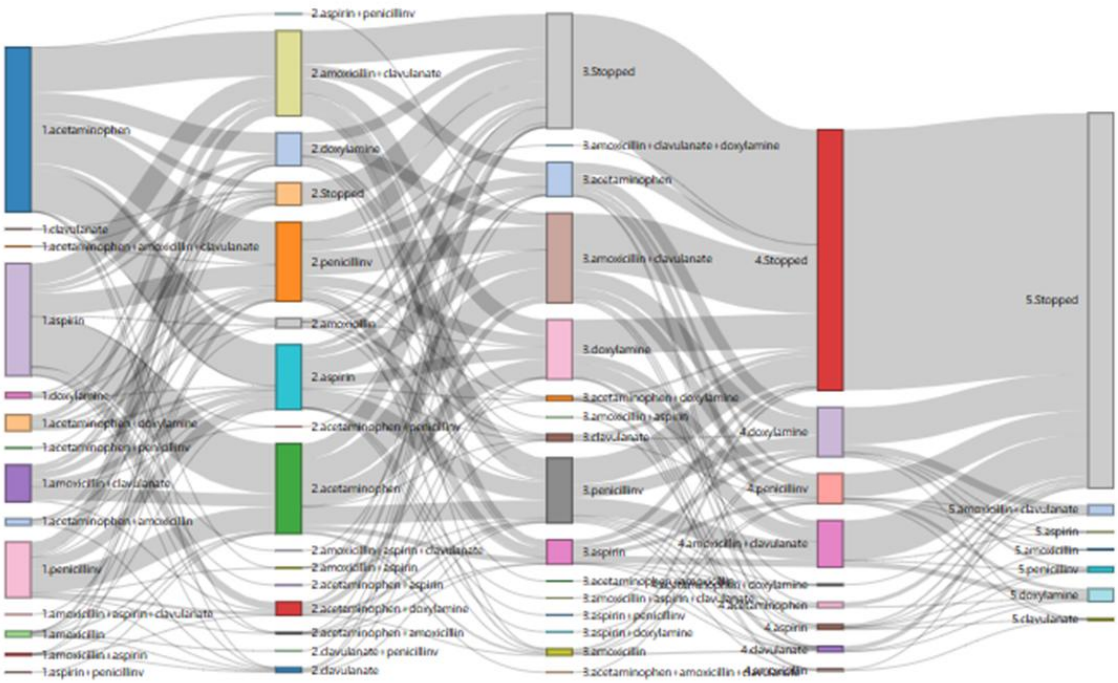

Figure 10. Sankey plot depicting treatment sequence over 10 years following HRT initiation in postmenopausal women in data source 1 *(objective 4)*.

Figures 9 and 10 are provided as mock visualisations to illustrate the intended graphical format. Similar Sunburst and Sankey plots will be generated for each participating data source and for age-stratified

analysis. The figures listed below (Figures 11–12) are not shown but will follow a similar structure and visual style, tailored to age-stratified analysis.

Figure 11. Sunburst plot depicting treatment patterns over 10 years following HRT initiation in postmenopausal women, stratified by age group in data source 1 *(objective 4)*.

Figure 12. Sankey plot depicting treatment sequence over 10 years following HRT initiation in postmenopausal women, stratified by age group, in data source 1 *(objective 4).*

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 data sources from four European countries, Denmark, Germany, Spain, and the United Kingdom, covering both registry-based and outpatient general practitioner care. The geographic coverage and longitudinal nature of these data sources enable assessment of treatment patterns over time in Europe. Use of the OMOP CDM ensures standardised data structuring, harmonised variables, and consistent cohort definitions, facilitating reproducible analyses across sources.

Limitations

This study will be informed by routinely collected healthcare data, which introduces several important considerations that may influence the interpretation of the findings. The results will reflect only the populations captured within the five data sources included in this study and may not be generalisable to other countries or healthcare systems.

Routinely collected electronic health records are primarily designed for clinical purposes rather than research, and as such, may contain incomplete, inconsistent, or variably recorded information. This affects the documentation of key variables such as comorbidities, symptoms, and clinical indications, which are essential for accurate patient-level characterisation.

Identification of postmenopausal women will be based on an age-based proxy across all data sources, due to inconsistent or absent documentation of menopause diagnoses. For early menopause, additional diagnostic codes, and conditions compatible with menopause will be used. While commonly used in observational research, this may lead to misclassification, particularly in cases of early menopause.

Data sources have information on the prescribing of medications, but information on whether prescriptions were redeemed or consumed is not available. Furthermore, some women may be treated in specialist centres, and prescriptions issued outside primary care practices might not be recorded in data sources, potentially resulting in underestimation of HRT use.

Indications for medication use are not recorded as such in the data sources. Therefore, medicines such as progestins may have been prescribed for conditions other than menopause management. Additionally, classification of HRT includes both medications specifically indicated for HRT and those that may serve as contraceptives. In postmenopausal women, such medications are considered HRT.

Estimation of treatment duration is also subject to limitations inherent to real-world data. Due to documentation gaps, end-of-treatment dates may be incomplete or lack precise end dates, requiring

methods aligned with OMOP CDM conventions. While this promotes consistency across sources, it may not fully capture true treatment variability and should be interpreted with caution.

The COVID-19 pandemic may have influenced healthcare access and prescribing practices, potentially affecting prevalence estimates of HRT use during the study period. Finally, while the intended study period spans from 1 January 1994 to 31 December 2024, the actual study period will vary across data sources due to differences in data availability and quality.

10. REFERENCES

1. Freeman, E.W. and K. Sherif, *Prevalence of hot flushes and night sweats around the world: a systematic review.* Climacteric, 2007. **10**(3): p. 197-214.
2. Kingsberg, S., et al., *Real-world evaluation of treatment utilization by women experiencing vasomotor symptoms associated with menopause in the United States and Europe: Findings from the REALISE study.* Maturitas, 2024. **189**: p. 108096.
3. Nappi, R.E., et al., *Prevalence and quality-of-life burden of vasomotor symptoms associated with menopause: A European cross-sectional survey.* Maturitas, 2023. **167**: p. 66-74.
4. Yang, L. and A. Toriola, *Menopausal Hormone Therapy Use Among Postmenopausal Women.* JAMA Health Forum, 2024. **5**(9): p. e243128.
5. Alsugeir, D., et al., *Hormone replacement therapy prescribing in menopausal women in the UK: a descriptive study.* BJGP Open, 2022. **6**(4).
6. Crandall, C.J., J.M. Mehta, and J.E. Manson, *Management of Menopausal Symptoms: A Review.* JAMA, 2023. **329**(5): p. 405-420.
7. Sprague, B.L., A. Trentham-Dietz, and K.A. Cronin, *A sustained decline in postmenopausal hormone use: results from the National Health and Nutrition Examination Survey, 1999-2010.* Obstet Gynecol, 2012. **120**(3): p. 595-603.
8. Davis, S.R. and R.J. Baber, *Treating menopause - MHT and beyond.* Nat Rev Endocrinol, 2022. **18**(8): p. 490-502.
9. Smith, T. and P. Batur, *Prescribing testosterone and DHEA: The role of androgens in women.* Cleve Clin J Med, 2021. **88**(1): p. 35-43.
10. Bakken, K., et al., *Menopausal hormone therapy and breast cancer risk: impact of different treatments. The European Prospective Investigation into Cancer and Nutrition.* Int J Cancer, 2011. **128**(1): p. 144-56.
11. Chlebowski, R.T., et al., *Association of Menopausal Hormone Therapy With Breast Cancer Incidence and Mortality During Long-term Follow-up of the Women's Health Initiative Randomized Clinical Trials.* JAMA, 2020. **324**(4): p. 369-380.
12. Fournier, A., F. Berrino, and F. Clavel-Chapelon, *Unequal risks for breast cancer associated with different hormone replacement therapies: results from the E3N cohort study.* Breast Cancer Res Treat, 2008. **107**(1): p. 103-11.
13. Tjonneland, A., et al., *Hormone replacement therapy in relation to breast carcinoma incidence rate ratios: a prospective Danish cohort study.* Cancer, 2004. **100**(11): p. 2328-37.
14. Markus, A.F., et al., *TreatmentPatterns: An R package to facilitate the standardized development and analysis of treatment patterns across disease domains.* Comput Methods Programs Biomed, 2022. **225**: p. 107081.
15. Rathmann, W., et al., *Basic characteristics and representativeness of the German Disease Analyzer database* Int J Clin Pharmacol Ther, 2018. **56**(10): p. 459-466.
16. Tanislav, C., et al., *No increased incidence of venous thrombosis or pulmonary embolism after SARS-CoV-2 vaccination in Germany.* Public Health, 2022. **207**: p. 14-18.
17. Ly, N.F., et al., *Impact of European Union Label Changes for Fluoroquinolone-Containing Medicinal Products for Systemic and Inhalation Use: Post-Referral Prescribing Trends.* Drug Saf, 2023. **46**(4): p. 405-416.

18. Sanchez-Santos, M.T., et al., *Data Resource Profile Update: CPRD GOLD.* Int J Epidemiol, 2025. **54**(4).

19. Carey, I.M., et al., *Prevalence of co-morbidity and history of recent infection in patients with neuromuscular disease: A cross-sectional analysis of United Kingdom primary care data.* PLoS One, 2023. **18**(3): p. e0282513.

20. Wigglesworth, S., et al., *The incidence and prevalence of epilepsy in the United Kingdom 2013-2018: A retrospective cohort study of UK primary care data.* Seizure, 2023. **105**: p. 37-42.

21. Fahmi, A., et al., *Combinations of medicines in patients with polypharmacy aged 65-100 in primary care: Large variability in risks of adverse drug related and emergency hospital admissions.* PLoS One, 2023. **18**(2): p. e0281466.

11. ANNEXES

ANNEX I. Data sources description

DATA SOURCE DESCRIPTION

Danish Data Health Registries (DK-DHR), Denmark

Denmark Danish health data is collected, stored, and managed in national health registers at the Danish Health Data Authority and covers the entire population, which makes it possible to study the development of diseases and their treatment over time. There are no gaps in terms of gender, age, and geography in Danish health data due to mandatory reporting on all patients from birth to death in all hospitals and medical clinics. Personal identification numbers enable linking of data across registers, so it captures data on all Danes throughout their lives, regardless of whether they have moved around the country. The high quality of Danish health data is attributed to standardisation, digitisation, and comprehensive documentation, which together enhance accuracy, consistency, and reliability, minimising potential for interpretation errors. The Danish Health Data Authority is responsible for the national health registers and for maintaining and developing standards and classifications in the Danish healthcare system. Legislation ensures balance between personal data protection and use. The current data release includes data on the entire Danish population of 5.9 million persons from 1995. It includes data from the following registries: The central Person Registry, The National Patient Registry, The Register of Pharmaceutical Sales, The National Cancer Register, The Cause of Death registry, the Laboratory Database (including coronavirus disease 2019 test results), and the Vaccination Registry (including COVID-19 vaccinations).

IQVIA Disease Analyzer Germany (IQVIA DA Germany), Germany

IQVIA Disease Analyzer (DA) Germany is a database of de-identified electronic medical records from specialised and general primary practices (GP) in Germany since 1992.[15] This dataset encompasses approximately 3% of all outpatient practices within Germany, ensuring a substantial representation of the national healthcare landscape. The sampling methods used for practice selection, taking into account physician's demographics, specialty focus, community size category and federal state location, were instrumental in constructing a database that accurately mirrors the diverse spectrum of healthcare providers in the country. Consequently, data within the IQVIA DA Germany database has been demonstrated to be representative of general and specialised practices throughout Germany.

The database contains demographics records, basic medical data, disease diagnoses according to the International Classification of Diseases, 10th revision (ICD-10), and prescription records. While the database partly records information on deaths and procedures, it currently does not support linkage with external data sources, and therefore, information on mortality is incomplete. Routine updates are conducted at regular intervals. Data quality is assessed based on several criteria, including completeness of information and correctness (e.g., linkage between diagnosis and prescriptions).

No registration or approval is required for drug utilisation studies. As previously demonstrated, IQVIA DA Germany is suitable for pharmacoepidemiologic and pharmacoeconomic studies.[16, 17]

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

BIFAP (http://www.bifap.org/index_EN.html) is a longitudinal population-based data source of medical patient records of the Spanish National Health Service (SNS) from 9 participating Regions throughout Spain out of the 17 Spanish Regions. The population currently included represents 36% of the total Spanish population. Spain has a SNS that provides universal access to health services through the Regional Healthcare Services. Primary care physicians (PCPs), both general practitioners and paediatricians, have a central role. They act as gatekeepers of the system and also exchange information with other levels of care to ensure the continuity of care. Most (98.9%) of the population is registered with a PCP, and, in addition, most drug prescriptions are written at the primary care level. BIFAP includes a collection of databases

linked at the individual patient level. The main one is the Primary care Database given the central role of PCPs in the SNS. Linked, there are additional important structural databases like the medicines dispensed at community pharmacies and the patients' hospital diagnosis at discharge. 7 out of the 9 regions have linkage to hospital data. However, hospital data is available for different time periods for each region. From 2014 onwards, linkage to hospital data is available for >68% of patients. Linkage to SARS-CoV-2 diagnostics tests and COVID-19 vaccination registries are also included. Additional databases are also linked for a subset of patients (hospital pharmacy, cause of death registry). BIFAP program is a non-profit program financed by the Spanish Agency of Medicines and Medical Devices (AEMPS), a government agency belonging to the Ministry of Health in collaboration with the regional health authorities.

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

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

Clinical Practice Research Datalink GOLD (CPRD GOLD), The United Kingdom

The Clinical Practice Research Datalink (CPRD) GOLD is a database of anonymised electronic health records (EHR) from General Practitioner (GP) clinics in the UK that use the Vision® software system for their management.[18] The source population registered with GPs responsible for non-emergency care and referrals. Participating GPs provide CPRD EHR for all registered patients who did not specifically request to opt out of data sharing. GOLD contains data from all four UK constituent countries, and the current regional distribution of its GP practices is 0.89% in England, 61.13% in Scotland, 26.41% in Wales, and 11.57% in Northern Ireland (January 2025).[18]

GOLD data include patient's demographic, biological measurements, clinical symptoms and diagnoses, referrals to specialists/hospitals and their outcomes, laboratory tests/results, and prescribed medications. GPs receive information about patient contacts with secondary care, but this information must be manually entered into the patient record and therefore, may be incomplete. GOLD has been assessed and found broadly representative of the UK general population in terms of age and sex.[18] GOLD has been widely used internationally for observational research to produce nearly 3,000 peer-reviewed publications, making GOLD the most influential UK clinical database so far.[19-21]

In terms of quality checks, the integrity, structure, and format of the data is reviewed. Collection-level validation ensures integrity by checking that data received from practices contain only expected data files and ensures that all data elements are of the correct type, length, and format. Duplicate records are identified and removed. Transformation-level validation checks for referential integrity between records ensure that there are no orphan records included in the database (for example, that all event records link to a patient), while research-quality-level validation covers the actual content of the data. CPRD provides a

patient-level data quality metric in the form of a binary ‘acceptability’ flag. This is based on recording and internal consistency of key variables, including date of birth, practice registration date and transfer out date.

ANNEX II. Additional information

DATA MANAGEMENT

Data management

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

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

Data storage and protection

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

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

The output files are stored in the DARWIN Remote Research Environment. These output files do not contain any data that allow identification of subjects included in the study. The RRE implements further security measures to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/20161 in the various member states.

QUALITY CONTROL

General data source quality control

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

Study specific quality control

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

allows the user to define a search strategy and using this will then query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, the *CohortDiagnostics* R package (https://github.com/OHDSI/CohortDiagnostics) will be run if needed to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error.

The study code will be based on four R packages to (1) characterise postmenopausal women initiating HRT treatment or not on HRT, (2) estimate prevalence of HRT use in postmenopausal women, (3) estimate treatment duration, and (4) describe treatment patterns using the OMOP CDM. 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® Coordination Centre (CC) upon completion of the study.

An interactive dashboard incorporating all the results (tables and figures) will be provided alongside the PDF report. The full set of underlying aggregated data used in the dashboard will also be made available if requested.

ANNEX III: List of stand-alone documents

Table S1. Preliminary list of conditions definitions indicating menopause.

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Menopause	Abnormal perimenopausal bleeding	4201098	-	SNOMED
Menopause	Climacteric arthritis	79109	-	SNOMED
Menopause	Climacteric arthritis of multiple sites	80806	-	SNOMED
Menopause	Climacteric arthritis of spine	4179506	-	SNOMED
Menopause	Climacteric arthritis of the ankle and/or foot	75341	-	SNOMED
Menopause	Climacteric arthritis of the hand	81376	-	SNOMED
Menopause	Climacteric arthritis of the pelvic region and thigh	75045	-	SNOMED
Menopause	Climacteric arthritis of the shoulder region	81938	-	SNOMED
Menopause	Climacteric flushing	4144760	-	SNOMED
Menopause	Dermatosis of menopause	4294304	-	SNOMED
Menopause	Endometrial cells, cytologically benign, in a postmenopausal woman	4052435	-	SNOMED
Menopause	Exacerbation of menopausal symptom	1340396	-	OMOP Extension
Menopause	Genitourinary syndrome of menopause	37209648	-	SNOMED
Menopause	Menopausal concentration lack	4112942	-	SNOMED
Menopause	Menopausal depression	4223090	-	SNOMED
Menopause	Menopausal flushing	4113666	-	SNOMED
Menopause	Menopausal headache	4113841	-	SNOMED
Menopause	Menopausal osteoporosis	4136988	-	SNOMED
Menopause	Menopausal problem	4286186	-	SNOMED
Menopause	Menopausal profile	4135533	-	SNOMED
Menopause	Menopausal sleeplessness	4113205	-	SNOMED
Menopause	Menopausal symptom	4071683	-	SNOMED
Menopause	Menopausal syndrome	439082	-	SNOMED
Menopause	Menopause caused by ionizing radiation	1076252	-	SNOMED
Menopause	Menopause due to and following radiotherapy	1076253	-	SNOMED
Menopause	Normal menopause	4034019	-	SNOMED
Menopause	Osteoporotic fracture due to postmenopausal osteoporosis	1245005	-	SNOMED
Menopause	Perimenopausal atrophic vaginitis	4255393	-	SNOMED
Menopause	Perimenopausal state	45757505	-	SNOMED
Menopause	Post-hysterectomy menopause	4175534	-	SNOMED
Menopause	Postartificial menopausal syndrome	440155	-	SNOMED

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Menopause	Postmenopausal androgenetic alopecia	4296346	-	SNOMED
Menopause	Postmenopausal atrophy of vocal cord	4048499	-	SNOMED
Menopause	Postmenopausal bleeding	195321	-	SNOMED
Menopause	Postmenopausal depression	37166876	-	SNOMED
Menopause	Postmenopausal flushing	4301021	-	SNOMED
Menopause	Postmenopausal frontal fibrosing alopecia	4296347	-	SNOMED
Menopause	Postmenopausal osteopenia	42536667	-	SNOMED
Menopause	Postmenopausal osteoporosis	4010333	-	SNOMED
Menopause	Postmenopausal postcoital bleeding	4213652	-	SNOMED
Menopause	Postmenopausal pruritus	4296644	-	SNOMED
Menopause	Postmenopausal state	4295261	-	SNOMED
Menopause	Postmenopausal urethral atrophy	4054881	-	SNOMED
Premature menopause	Postsurgical menopause	4154697	-	SNOMED

Table S2. Preliminary list of conditions definitions indicating early menopause.

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Premature menopause	Primary ovarian failure	4279913	-	SNOMED
Premature menopause	Premature menopause	198715	-	SNOMED
Premature menopause	Postsurgical menopause	4154697	-	SNOMED
Premature menopause	Laparoscopic bilateral oophorectomy	43531553	-	SNOMED
Premature menopause	Total hysterectomy with removal of both tubes and ovaries	4312859	-	SNOMED
Premature menopause	Bilateral acquired absence of ovary	36716936	-	SNOMED

Table S3. Preliminary list of measurements/conditions/procedures.

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
BMI	Body mass index	4245997	4087497	SNOMED
BMI	Finding of body mass index	4103471	-	SNOMED
BMI	Measurement of body mass index	44783982	-	SNOMED
BMI	Body mass index (BMI) [Ratio]	3038553	-	LOINC
Smoking	Smoking-related terms	40654161	-	LOINC
Smoking	Cigarette smoker	903657	-	OMOP Extension

Phenotype	Concept name	Concept id (including descendants)	Exclude concept id	Vocabulary
Smoking	Tobacco or its derivatives user	903654	-	OMOP Extension
Cardiovascular disease	Disorder of cardiovascular system	134057	-	SNOMED
Hypertension	Essential hypertension	320128	-	SNOMED
Diabetes	Diabetes mellitus	201820	-	SNOMED
Dyslipidaemia	Dyslipidaemia	4159131	-	SNOMED
Metabolic syndrome	Metabolic syndrome X	436940	-	SNOMED
Osteoporosis	Osteoporosis	80502	-	SNOMED
Osteoporosis	Osteoporotic fracture	40480160	-	SNOMED
Hysterectomised or intact uterus	Hysterectomy	4127886	-	SNOMED

Table S4. Preliminary list of medicines definitions.

Class	WHO ATC	Concept name	Ingredient Concept ID	Include descendants
Natural and semisynthetic oestrogens	G03CA	Estradiol	1548195	Yes
Natural and semisynthetic oestrogens	G03CA	Estriol	19049038	Yes
Natural and semisynthetic oestrogens	G03CA	Conjugated equine oestrogens	1549080	Yes
Natural and semisynthetic oestrogens	G03CA	Esterified oestrogens	1551673	Yes
Natural and semisynthetic oestrogens	G03CA	Chlorotrianisene	19092684	Yes
Natural and semisynthetic oestrogens	G03CA	Ethinylestradiol	1549786	Yes
Natural and semisynthetic oestrogens	G03CA	Estrone	1549254	Yes
Natural and semisynthetic oestrogens	G03CA	Promestriene	19029079	Yes
Synthetic oestrogens	G03CB	Dienestrol	925102	Yes
Synthetic oestrogens	G03CB	Methallenestril	36863204	Yes
Synthetic oestrogens	G03CB	Moxestrol	36852976	Yes
Synthetic oestrogens	G03CB	Diethylstilbestrol	1525866	Yes
Other oestrogens	G03CX	Tibolone	19041933	Yes
Progesterones	G03D	Micronised progesterone	1552310	Yes
Progesterones	G03D	Dydrogesterone	19040060	Yes
Progesterones	G03D	Norethisterone	1521369	Yes
Progesterones	G03D	Levonorgestrel	1589505	Yes
Progesterones	G03D	Medroxyprogesterone acetate	1500211	Yes
Progesterones	G03D	Lynestrenol	19092358	Yes
Progesterones	G03D	Megestrol / Medrogestone	1300978/19000207	Yes
Progesterones	G03D	Nomegestrol	19015533	Yes

ANNEX IV: ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

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

Study title: DARWIN EU® - Treatment patterns in postmenopausal women in European countries: A real-world observational study

EU PAS Register® number: EUPAS1000000821
Study reference number (if applicable): P4-C1-013

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

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	
2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	7
2.1.2 The objective(s) of the study?	☒	☐	☐	
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☒	☐	☐	

Comments:

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

Section 3: Study design		**Yes**	**No**	**N/A**	**Section Number**
3.1	Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	8.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	8.2
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	8.8
3.4	Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☒	
3.5	Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

Comments:

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

Comments:

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	8.6.1
5.2	Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☐	☒	
5.3	Is exposure categorised according to time windows?	☐	☐	☒	
5.4	Is intensity of exposure addressed? (e.g. dose, duration)	☒	☐	☐	8.8.3
5.5	Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
5.6	Is (are) (an) appropriate comparator(s) identified?	☐	☐	☒	

Comments:

Section 6: Outcome definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
6.1	Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	8.6.2
6.2	Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.6.2
6.3	Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☐	☒	
6.4	Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

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

Comments:

Section 8: Effect measure modification		**Yes**	**No**	**N/A**	**Section Number**
8.1	Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☐	☐	☒	

Comments:

Section 9: Data sources		**Yes**	**No**	**N/A**	**Section Number**
9.1	Does the protocol describe the data source(s) used in the study for the ascertainment of:				
	9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☒	☐	☐	8.2, 8.6
	9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	8.2, 8.6
	9.1.3 Covariates and other characteristics?	☒	☐	☐	8.2, 8.6
9.2	Does the protocol describe the information available from the data source(s) on:				
	9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☒	☐	☐	8.2, 8.6
	9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	8.2, 8.6
	9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	8.2, 8.6
9.3	Is a coding system described for:				
	9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	8.2, 8.6
	9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	8.2, 8.6
	9.3.3 Covariates and other characteristics?	☒	☐	☐	8.2, 8.6
9.4	Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☒	

Comments:

Section 10: Analysis plan		**Yes**	**No**	**N/A**	**Section Number**
10.1	Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.8
10.2	Is study size and/or statistical precision estimated?	☒	☐	☐	8.8
10.3	Are descriptive analyses included?	☒	☐	☐	8.8
10.4	Are stratified analyses included?	☒	☐	☐	8.8
10.5	Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6	Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	
10.7	Does the plan describe methods for handling missing data?	☐	☐	☒	
10.8	Are relevant sensitivity analyses described?	☒	☐	☐	8.8

Comments:

Section 11: Data management and quality control		**Yes**	**No**	**N/A**	**Section Number**
11.1	Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	11
11.2	Are methods of quality assurance described?	☒	☐	☐	11
11.3	Is there a system in place for independent review of study results?	☐	☐	☒	

Comments:

Section 12: Limitations		**Yes**	**No**	**N/A**	**Section Number**
12.1	Does the protocol discuss the impact on the study results of:				
	12.1.1 Selection bias?	☒	☐	☐	
	12.1.2 Information bias?	☒	☐	☐	9
	12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☐	☐	☒	
12.2	Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	8.2, 8.7

Comments:

Section 13: Ethical/data protection issues		**Yes**	**No**	**N/A**	**Section Number**
13.1	Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	8.2
13.2	Has any outcome of an ethical review procedure been addressed?	☒	☐	☐	8.2
13.3	Have data protection requirements been described?	☒	☐	☐	11

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)?	☒	☐	☐	11
15.2	Are plans described for disseminating study results externally, including publication?	☒	☐	☐	11

Comments:

Name of the main author of the protocol: Dina Vojinovic

Date: 20/01/2026

Signature: 

ANNEX V: 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 Centre in Rotterdam, the Netherlands.

Data Access

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

Data Quality Framework

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

Data Source

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

DARWIN EU®

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

EMA (European Medicines Agency)

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

Evidence Generation

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

Federated Network

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

GDPR (General Data Protection Regulation)

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

Health Technology Assessment (HTA)

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

Metadata

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

Off-the-Shelf Studies (OTS)

Studies for which a standard protocol per study/analysis type and standardised analytics may be developed and applied or adapted, typically relating to a descriptive research question. This includes studies on disease epidemiology, for example, the estimation of the prevalence or incidence of health outcomes in defined time periods and population groups, or drug utilisation studies at the population or patient level.

OHDSI (Observational Health Data Sciences and Informatics)

An open-science collaborative community that develops tools and standards (including the OMOP CDM) to enable large-scale analytics of observational health data. OHDSI provides the technical and scientific foundation for DARWIN EU®'s analytical ecosystem.

Patient-Level Data

Data related to individuals, de-identified, used for longitudinal or detailed analyses.

OMOP (Observational Medical Outcomes Partnership)

A common data model (CDM) that standardises the structure and content of observational healthcare data, enabling systematic analysis across disparate datasets. DARWIN EU® uses the OMOP CDM to ensure interoperability and consistency in real-world evidence generation.

Real-World Data (RWD)

Data relating to individual health status or healthcare delivery that is collected from routine clinical practice rather than from randomised controlled trials.

Real-World Evidence (RWE)

Clinical evidence derived from the analysis of RWD, used to inform decisions by regulators, payers, or clinicians.

Regulatory Decision-Making

The process by which authorities like EMA assess data to authorise, monitor, or modify the use of medicines in the EU.

Routine Repeated Studies (RR)

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

Study Protocol

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

Very Complex Studies (C4)

Studies which cannot rely only on electronic healthcare 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.