Study Protocol

P4-C2-013, P4-C2-014, and P4-C2-021

DARWIN EU® - Determinants for use of GLP1 receptor agonists – a drug utilisation study

27/02/2026

Version 6.0

Authors: Marta Pineda Moncusí,
Annika Jödicke, Martí Català Sabaté

Public

CONTENTS

<table>
<tr><td>Study title[1]</td><td>DARWIN EU® - Determinants for use of GLP1 receptor agonists – a drug utilisation study</td></tr>
<tr><td>Protocol version</td><td>V5.0</td></tr>
<tr><td>Date</td><td>10/02/2026</td></tr>
<tr><td>EUPAS number</td><td>EUPAS1000000828</td></tr>
<tr><td>Active substance</td><td>Use of the following GLP-1 RA ingredients:
<table>
<tr><th>Ingredient name</th><th>ATC code</th><th>Concept ID</th></tr>
<tr><td>Exenatide</td><td>A10BJ01</td><td>1583722</td></tr>
<tr><td>Liraglutide</td><td>A10BJ02</td><td>40170911</td></tr>
<tr><td>Lixisenatide</td><td>A10BJ03</td><td>44506754</td></tr>
<tr><td>Dulaglutide</td><td>A10BJ05</td><td>45774435</td></tr>
<tr><td>Semaglutide</td><td>A10BJ06</td><td>793143</td></tr>
<tr><td>Tirzepatide</td><td>A10BX16</td><td>779705</td></tr>
</table></td></tr>
<tr><td>Medicinal product</td><td>N/A</td></tr>
<tr><td>Research question and objectives</td><td><u>Research question</u>

This study is a routinely repeated studies of a previous DARWIN EU® study (P3-C1-008)[1] (with substantial adaptations) to answer the research question “What are the patterns of use of GLP-1 RA in 2015 – 2025?”.

<u>Objectives</u>

The study will comprise the following objectives:

1) To determine the incidence and prevalence of prescriptions of the GLP-1 RA medicines (overall, by ingredient, by pre-specified brand) during the last 10 years of available data, stratified separately by age, sex, indication (pre-defined presence or absence of diagnosis of type 2 diabetes mellitus and/or obesity) [only for incidence], and prescriber speciality (where available) [only for incidence], and by calendar month in each of the data sources.

2) To characterise new drug users of GLP-1 RA medicines (overall and stratified by ingredient, by brand, by indication cohorts [pre-defined presence or absence of diagnosis of type 2 diabetes mellitus and/or diagnosis of obesity) by age, sex, initial dose, cumulative dose, and a list of prespecified indications/comorbidities and related co-medications for each data source. Characterisation will be done over the whole study period and by calendar year.

3) To describe GLP-1 RA switching of substances (exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide, or tirzepatide cohorts) among new GLP-1 RA users for each data source [overall study period].

4) To describe GLP-1 RA within-substance switching of strength (limited to pre-defined cohorts of strength-levels, with brands when available, for liraglutide, dulaglutide, semaglutide, and tirzepatide) among new users of the respective medicine in each data source [overall study period].

5) To describe GLP-1 RA switching of substances (exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide, or tirzepatide cohorts) among new GLP-1 RA users for each data source between 2015–2020 and annually between 2021 and 2025.

6) To describe GLP-1 RA within-substance switching of strength (limited to pre-defined cohorts of strength-levels, with brands when available, for liraglutide, dulaglutide, semaglutide, and tirzepatide) among new users of the respective medicine in each data source between 2015–2020 and annually between 2021 and 2025.</td></tr>
</table>

Countries of study	Belgium, Croatia, Denmark, Finland, Germany, the Netherlands, Norway, Spain, Sweden, and the United Kingdom.
Authors	Marta Pineda Moncusí, m.pinedamoncusi@darwin-eu.org Annika Jödicke, a.jodicke@darwin-eu.org Martí Català Sabaté, m.catalasabate@darwin-eu.org

[1]*This is a routinely repeated study from P3-C1-008 with EUPAS1000000223.*

1. LIST OF ABBREVIATIONS

Acronyms/term	Description
ATC	Anatomical Therapeutic Chemical
ADHD	Attention-Deficit/Hyperactivity Disorder
AIDS	Acquired Immunodeficiency Syndrome
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CC	Coordination Centre
CDM	Common Data Model
COPD	Chronic Obstructive Pulmonary Disease
CPRD GOLD	Clinical Practice Research Datalink GOLD
CVD	Cardiovascular Disease
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
T2DM	Type 2 diabetes mellitus
DOI	Declaration of Interests
DPP-4 inhibitors	Dipeptidyl Peptidase 4 inhibitors
DQD	Data Quality Dashboard
DRE	Digital Research Environment
DUS	Drug Utilization Study
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
FinOMOP-THL	Finnish Care Register for Health Care
GDPR	General Data Protection Regulation
GLP-1 RA	Glucagon-like peptide-1 receptor agonists
GIP	Glucose-dependent insulinotropic polypeptide
GP	General Practitioner
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage
ICD	International Classification of Diseases
InGef RDB	InGef – Institute for Applied Health Research Berlin GmbH, RDB – Research Database
IP	Inpatient
IPCI	Integrated Primary Care Information Project
IQR	Interquartile Range
IQVIA DA Germany	IQVIA Disease Analyzer Germany
IQVIA LPD Belgium	IQVIA Longitudinal Patient Database Belgium

Acronyms/term	Description
IRB	Institutional Review Board
LADA	Latent Autoimmune Diabetes in Adults
MACE	Major Adverse Cardiovascular Event
MASLD	Metabolic Dysfunction-Associated Steatotic Liver
NA	Not applicable
NAJS	Croatian National Public Health Information System
NLHR	Norwegian Linked Health Registry data
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
PCOS	Polycystic Ovary Syndrome
RxNorm	Medical prescription normalized
SGLT2 inhibitors	Sodium-Glucose co-Transporter 2 inhibitors
SIDIAP	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària
SNOMED	Systematized Nomenclature of Medicine
WHO	World Health Organisation

2. TITLE

DARWIN EU® - Determinants for use of GLP1 receptor agonists – a drug utilisation study.

3. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Marta Pineda Moncusí	University of Oxford
Data Scientist	Xihang Chen Martí Català Sabaté	University of Oxford University of Oxford
Epidemiologists	Annika Jödicke Martí Català Sabaté	University of Oxford University of Oxford
Clinical Domain Expert	Annika Jödicke Anna Saura Lazaro	University of Oxford University of Oxford
Study Manager	Natasha Yefimenko	Erasmus Medical Centre
Data Partner* P4-C2-013	**Names**	**Organisation**
CPRD GOLD	Antonella Delmestri	University of Oxford
DK-DHR	Elvira Bräuner (Lead epidemiologist) Susanne Bruun (Statistical programmer)	Danish Medicines Agency (DKMA)
IPCI	Katia Verhamme Mees Mosseveld	Erasmus University Medical Center
IQVIA LPD Belgium	Ellen Gerritsen Dina Vojinovic Akram Mendez Gargi Jadhav	IQVIA
IQVIA DA Germany		
SIDIAP	Anna Palomar Cros Agustina Giuliodori Picco Irene López Sánchez Laura Granés González	IDIAPJGOL
Data Partner* P4-C2-014	**Names**	**Organisation**
FinOMOP-THL	Laura Salonen Toni Lehtonen Petteri Hovi Terhi Kilpi Tuomo Susi	Finnish Institute for Health and Welfare (THL)
HI-SPEED	Rickard Ljung Marcel Ballin Mats Talbäck	Swedish Medical Products Agency - Gothenburg University (SMPA-GU)

NAJS	Jakov Vukovic Tamara Buble Jelena Dimnjakovic Zeljka Drausnik Danijela Fustin Antea Jezidzic Marko Cavlina	Croatian Institute of Public Health
InGef RDB	Alexander Harms Raeleesha Norris Annika Vivirito	InGef - Institute for Applied Health Research
NLHR	Hedvig Nordeng Saeed hayati Marleen van Gelder	University of Oslo

*Data partners do not have an investigator role. Data partners execute code at their data source, review and approve their results. Data partners participating in **P4-C2-021** (i.e., Objectives 5 and 6) will be selected from the above, subject to feasibility of execution.

4. ABSTRACT

Title

DARWIN EU® - Determinants for use of GLP1 receptor agonists – a drug utilisation study

Rationale and background

A shortage of medicines containing Glucagon-Like Peptide-1 receptor agonists (GLP-1 RA) has been affecting European Union (EU) Member States since 2022. The medicines belonging to the class of GLP-1 RA are either authorised for the treatment of diabetes or authorised for weight management in patients diagnosed with obesity, with the exception of Mounjaro (tirzepatide), a glucose-dependent insulinotropic polypeptide (GIP) and GLP-1 RA that is authorised for both indications.

Increased demand for these medicines has contributed to this shortage in addition to other causes, e.g., capacity constraints. Excessive off-label use for cosmetic weight loss of some of these medicines has raised concerns. This relates to use for weight management in people without obesity or people with overweight who do not have weight related health problems. This use has been mentioned frequently in the news and social media and is exacerbating existing shortages with serious consequences for public health.(1)

This study aims to provide an overview of the characteristics of patients prescribed with GLP-1 RA and to describe the pattern of use, including switching between GLP-1 RA substances and to other antidiabetics, as well as switching between selected brands. The study will span over 10 years, allowing for the assessment of potential changes in incidence of use, user characteristics, and treatment patterns. This will help contextualise what determinants might be driving the demand for GLP-1 RA in relation to the observed shortage of medicines, including exploring comparative trends of prescription of other medicinal products used in diabetes and for weight management, as well as patterns of off-label use. (2)

Research question and objectives

This study is a routinely repeated study of a previous DARWIN EU® study (P3-C1-008, DARWIN EU® - Drug Utilisation Study on GLP-1 Receptor Agonists), with substantial adaptations described in **section 5**. Amendments and updates to answer the research question “What are the patterns of use of GLP-1 RA in 2015 – 2025?”.

Objectives

The study will comprise the following objectives:

1) To determine the incidence and prevalence of prescriptions of the GLP-1 RA medicines (overall, by ingredient, by pre-specified brand) during the last 10 years of available data, stratified separately by age, sex, indication (pre-defined presence or absence of diagnosis of type 2 diabetes mellitus and/or diagnosis of obesity) [only for incidence], and prescriber speciality (where available) [only for incidence], and by calendar month in each of the data sources.
2) To characterise new drug users of GLP-1 RA medicines (overall and stratified by ingredient, by brand, by indication cohorts [pre-defined presence or absence of diagnosis of type 2 diabetes mellitus and/or diagnosis of obesity) by age, sex, initial dose, cumulative dose, and a list of prespecified indications/comorbidities and related co-medications for each data source. Characterisation will be done over the whole study period and by calendar year.
3) To describe GLP-1 RA switching of substances (exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide or tirzepatide cohorts) among new GLP-1 RA users for each data source [overall study period].
4) To describe GLP-1 RA within-substance switching of strength (limited to pre-defined cohorts of strength-levels, with brands when available, for liraglutide, dulaglutide, semaglutide, and tirzepatide) among new users of the respective medicine in each data source [overall study period].

5) To describe GLP-1 RA switching of substances (exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide or tirzepatide cohorts) among new GLP-1 RA users for each data source between 2015–2020 and annually between 2021 and 2025.
6) To describe GLP-1 RA within-substance switching of strength (limited to pre-defined cohorts of strength-levels, with brands when available, for liraglutide, dulaglutide, semaglutide, and tirzepatide where strength-level cohorts can be reliably computed) among new users of the respective medicine in each data source between 2015–2020 and annually between 2021 and 2025.

Methods

Study design

- Population-level drug utilisation study (Objective 1. Population-level cohort on GLP-1 RA use)
- Patient-level drug utilisation study (Objective 2. New drug/s user cohort on GLP-1 RA)
- Patient-level characterisation (Objectives 3 and 5. Cohort analyses on new GLP-1 RA users)
- Patient-level characterisation (Objectives 4 and 6. Cohort analyses on new liraglutide, dulaglutide, semaglutide, and tirzepatide users)

Population

The study population will include all individuals present in the database during the study period: 01/01/2015 up until the latest data available for each data partner. For incident users (a.k.a., new users), a minimum of 365 days of data availability and no use of the respective GLP-1 RA during the previous 365 days will be required. In case of objectives 3 and 5, it will be restricted to no use of any GLP-1 RA ingredient during the previous 365 days.

Variables

Exposure:

Overall use of GLP-1 RA, by ingredient (exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide, and tirzepatide), and by pre-specified brands (Wegovy, Ozempic, Rybelsus, Trulicity, Victoza, Saxenda, and Mounjaro).

Cohorts of pre-defined strength-levels within GLP-1 RA ingredients (restricted to liraglutide, dulaglutide, semaglutide, and tirzepatide) and within pre-specified brands (Wegovy, Ozempic, Rybelsus, Trulicity, Victoza, Saxenda, and Mounjaro). Other brands will be included in a 'Other brands/no brand recorded' cohort.

Other second-line antidiabetics at ATC3 or ATC4 level class, and insulin ATC3 level class, for the characterisation of GLP-1 RA switching of substances.

Relevant covariates:

For stratification: overall, age, sex, indication cohorts (type 2 diabetes mellitus /obesity), prescriber speciality (where available), calendar month [incidence/prevalence], calendar year.

For description: age, sex, Body Mass Index (BMI), comorbidities, related co-medications, initial and cumulative dose, treatment duration, proportion of first prescription records mapped to ingredient level in GLP-1 RA users. Additionally, if possible: ethnicity, geographic stratification, and post-partum for women during the last 12 months before first GLP-1 RA use.

Data sources

1. Belgium: IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)
2. Croatia: Croatian National Public Health Information System (NAJS)

3. Denmark: Danish Data Health Registries (DK-DHR)
4. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
5. Germany: InGef Research Database (InGef RDB)
6. Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)
7. The Netherlands: Integrated Primary Care Information (IPCI)
8. Norway: Norwegian Linked Health Registry data (NLHR)
9. Spain: The Information System for Research on Primary Care (SIDIAP)
10. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
11. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Study size

No sample size has been calculated, as this is an exploratory study which will not test a specific hypothesis. Based on a preliminary feasibility assessment, the expected number of persons counts for the GLP-1 RA drugs in the data sources included in this study range from <1,000 (for lixisenatide in IQVIA LPD Belgium, DK-DHR, FinOMOP-THL, HI-SPEED, InGef RDB, IPCI, and NLHR) to >300,000 (for semaglutide in DK-DHR).

Statistical analysis

- Population-level drug utilisation: Monthly incidence rates and monthly point prevalence estimates, both with 95% confidence interval.
- Patient-level drug utilisation: New users of GLP-1 RA will be characterised by the relevant covariates listed above for the overall study period and by calendar year. The characterisation will be calculated for overall GLP-1 RA use, by ingredient, and by pre-specified brand.
- Patient-level characterisation of GLP-1 RA switching of substances and of GLP-1 RA within-substance switching of strength [i.e., Objectives 3-6]: Switching will be reported using Sankey diagram. Time-to-switch will be based on survival analysis. We will report median time-to-switch, (based on median survival: time where 50% of the patients in the group had switch to another treatment/strength), average time-on-treatment (based on restricted mean survival: average time patients are alive and still had not changed the treatment/strength over the whole follow-up period [10 years]), and Kaplan-Meier plots. Analysis will be reported for the overall study period [Objectives 3 and 4]; and between 2015–2020, and annually between 2021 and 2025 [Objectives 5 and 6]. Objectives 5 and 6 will be carried out subject to feasibility of execution.

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

5. AMENDMENTS AND UPDATES

This is a routinely repeated study from the P3-C1-008 study, which was registered in the catalogue (https://catalogues.ema.europa.eu/node/4124/administrative-details) with the EUPAS register number EUPAS1000000223. The study code from P3-C1-008 is available in the GitHub repository: https://github.com/darwin-eu-studies/P3-C1-008-GLP-1/

The following updates are the changes included in the routinely repeated analysis P4-C2-013 and P4-C2-014 when compared with P3-C1-008:

Number	Previous approved version of the protocol	Date	Section of study protocol	Amendment or Update	Reason
P3-C1-008	V3	17/10/2025	All protocol	Update of the protocol to fit the new protocol template.	To align with the new protocol format, in which sections were simplified and standardised text was updated. Any change in the methodological content of the study is described in the cells below.
P3-C1-008	V3	17/10/2025	**RESEARCH QUESTION AND OBJECTIVES;** **9.6.1 Exposure**	Removing prior objective 2. To determine the incidence and prevalence of prescriptions of medicines used for weight loss that help contextualise exposure to GLP-1 RA,	Results on other treatments for diabetes and weight management were limited. Thus, they were excluded.
P3-C1-008	V3	17/10/2025	**RESEARCH QUESTION AND OBJECTIVES**	Prior objective 1 (To determine incidence and prevalence of GP-1 RA use and to characterise new GLp-1 RA users) had been split in Objective 1 and 2.	Clarification of the text
P3-C1-008	V3	17/10/2025	**RESEARCH QUESTION AND OBJECTIVES**; **RESEARCH METHODS**	Objective 1. Incidence/Prevalence will be calculated overall, by ingredient and by pre-specified brand (not only by ingredient).	To provide further understanding of drug use patterns.
P3-C1-008	V3	17/10/2025	**RESEARCH QUESTION AND OBJECTIVES**; **RESEARCH METHODS**	Objective 2. Characterisation has been expanded by: - adding overall, by brand, and calendar year stratifications. -adding additional patient characteristics such as ethnicity (if available), comorbidities, and co-medications.	To provide further understanding of patients' characteristics.
P3-C1-008	V3	17/10/2025	**RESEARCH QUESTION AND OBJECTIVES**;	Addition of objectives 3 and 4 Characterisation of GLP-1 RA utilisation patterns (including 'Within GLP-1 RA switching of	Interest in understanding the treatment pattern of new GLP-1 RA users.

Number	Previous approved version of the protocol	Date	Section of study protocol	Amendment or Update	Reason
			RESEARCH METHODS	substances', 'Within-substance switching of strength' and time-to-switch), for the overall study period.	
P3-C1-008	V3	17/10/2025	RESEARCH METHODS	Sensitivity analysis to characterise the use of individual strengths has been replaced by the 'Within-substance switching of strength' analysis from Objective 4.	To provide further understanding of dose escalation.
P3-C1-008	V3	09/01/2026	RESEARCH QUESTION AND OBJECTIVES; RESEARCH METHODS	Inclusion of Objectives 5 and 6, where we analyse Objectives 3 and 4, respectively, for 2015–2020 and annually between 2021-2025. Objective 5 and 6 will be part of the routinely repeated analysis P4-C2-021.	To provide further understanding of GLP-1 RA shortage effect on switching.

6. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	February 2026
Creation of Analytical code	February-March 2026
Execution of Analytical Code on the data	April 2026
Draft Study Report	May 2026
Final Study Report	To be confirmed by EMA

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

7. RATIONALE AND BACKGROUND

A shortage of medicines containing Glucagon-Like Peptide-1 receptor agonists (GLP-1 RA) has been affecting European Union (EU) Member States since 2022.(3, 4) The medicines belonging to the class of GLP-1 RA are either authorised for the treatment of diabetes or authorised for weight management in patients diagnosed with obesity,(5, 6) with the exception of Mounjaro (tirzepatide), a glucose-dependent insulinotropic polypeptide (GIP) and GLP-1 RA that is authorised for both indications.(7)

Increased demand for these medicines has contributed to this shortage in addition to other causes, e.g., capacity constraints. Excessive off-label use for cosmetic weight loss of some of these medicines has raised concerns. This relates to use for weight management in people without obesity or people with overweight who do not have weight related health problems. This use has been mentioned frequently in the news and social media and is exacerbating existing shortages with serious consequences for public health.(1)

This study aims to provide an overview of the characteristics of patients prescribed with a GLP-1 RAs and to describe the pattern of use, including switching between substances and to other antidiabetics, as well as switching between selected brands. The study will span over up to 10 years, allowing for the assessment of potential changes in incidence of use, user characteristics, and treatment patterns.

This protocol contains a routinely repeated study of a previous DARWIN EU® study (EUPAS1000000223) focused on providing an overview of the characteristics of patients prescribed with a GLP-1 RA medicinal product and how these have changed over the past ten years.(2) This study is now being repeated to expand on the knowledge of treatment patterns of new GLP-1 RA users, including switching between GLP-1 RA substances and to other antidiabetics as well as switching between selected brands, and to include more recent data and additional data sources. This will help contextualise what determinants might be driving the demand for GLP-1 RA in relation to the observed shortage of medicines, including exploring comparative trends of prescription of other medicinal products used in diabetes and for weight management, as well as patterns of off-label use. (2)

8. RESEARCH QUESTION AND OBJECTIVES

Research questions

This study is a routinely repeated studies of a previous DARWIN EU® study (P3-C1-008, with substantial adaptations) to answer the research question “What are the patterns of use of GLP-1 RA in 2015 – 2025?”.

Research objectives

The study will comprise the following objectives:

1) To determine the incidence and prevalence of prescriptions of the GLP-1 RA medicines (overall, by ingredient, by brand [pre-defined]) during the last 10 years of available data, stratified separately

by age, sex, indications (type 2 diabetes mellitus /obesity) [only for incidence] and prescriber speciality (where available) [only for incidence], and by calendar month in each of the data sources.

2) To characterise new drug users of GLP-1 RA medicines (overall and stratified by ingredient, by brand, by indication cohorts [pre-defined presence or absence of diagnosis of type 2 diabetes mellitus and/or diagnosis of obesity indications]) by age, sex, initial dose, cumulative dose, and a list of prespecified indications/comorbidities and related co-medications for each data source. Characterisation will be done over the whole study period and by calendar year.
3) To describe GLP-1 RA switching of substances (exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide, or tirzepatide cohorts) among new GLP-1 RA users for each data source [overall study period].
4) To describe GLP-1 RA within-substance switching of strength (limited to pre-defined cohorts of strength-levels, with brands when available, for liraglutide, dulaglutide, semaglutide, and tirzepatide) among new users of the respective medicine in each data source [overall study period].
5) To describe GLP-1 RA switching of substances (exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide or tirzepatide cohorts) among new GLP-1 RA users for each data source between 2015–2020 and annually between 2021 and 2025.
6) To describe GLP-1 RA within-substance switching of strength (limited to pre-defined cohorts of strength-levels, with brands when available, for liraglutide, dulaglutide, semaglutide, and tirzepatide) among new users of the respective medicine in each data source between 2015–2020 and annually between 2021 and 2025.

9. RESEARCH METHODS

9.1. Study design

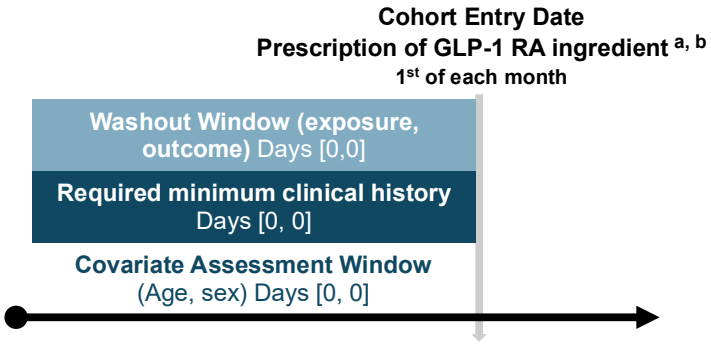
A cohort study will be conducted using routinely collected health data from 11 data sources from 10 countries across Europe. The study will comprise of:

- A population-level drug utilisation study will be conducted to address objective 1, determining the incidence and prevalence of prescriptions of the GLP-1 RA medicines.
- A patient-level drug utilisation study will be conducted to address objective 2, characterising new GLP-1 RA users.
- A patient-level characterisation will be conducted to address objectives 3 and 5, characterising the GLP-1 RA utilisation pattern in new users of GLP-1 RA medicines with no prior use of any other GLP-1 RA ingredient in the last 365 days.
- A patient-level characterisation will be conducted to address objectives 4 and 6, characterising the GLP-1 RA utilisation pattern in new liraglutide, semaglutide, dulaglutide, and tirzepatide users.

Objectives 5 and 6 will be carried out subject to feasibility of execution.

A description of the study design is depicted in **Figure 1**. Further details and visuals of the outcomes of the study that will be generated are described in **Annex I**.

A. Prevalent patients (persistent users) [Obj. 1 – Point prevalence]



B. Incident patients (new users of the medicine of interest) [Obj. 1 – Incidence, Obj. 2– 6]

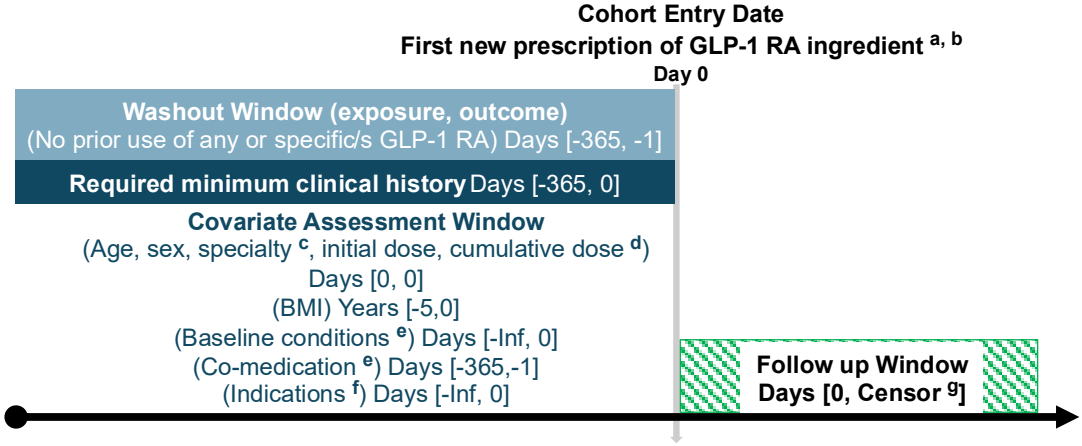


Figure 1. Graphical depiction of the study design.

a. GLP-1 RA ingredients: exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide, and tirzepatide.
b. Treatment episodes are defined by date of prescription/dispensation and days supply with a stockpiling algorithm if a new dispensation occurs before the end of days supply. Gaps of less than 90 days between end of days supply and next dispensation are bridged. Thirty days are added to the last dispensation's days supply in an exposure episode.
c. Where available (e.g., GP, endocrinologist).
d. Cumulative dose on first continuous treatment episode.
e. Baseline conditions and related co-medications included are fully described in **section 9.6.3**.
f. Indications for GLP-1 RA use included: diagnosis of type 2 diabetes mellitus and obesity.
g. Earliest of: loss to follow-up, death, or the end of observation period (the latest available data). In Objectives 3 and 5, follow-up will additionally be censored when switching from a GLP-1 RA drug to a non-GLP-1 RA drug listed in **Table 4**, or discontinuation of the GLP-1 RA treatment not followed by any drug listed in **Table 4**. In Objectives 4 and 6, follow-up will be additionally censored at discontinuation of the specific GLP-1 receptor agonist ingredient.

BMI = Body mass index.
GLP-1 RA = Glucagon-Like Peptide-1 receptor agonists.

9.2. Study setting and data sources

This study will be conducted using routinely collected data from 11 data sources in the DARWIN EU® network of data partners from 10 European countries. All data were a priori mapped to the OMOP CDM.

Data sources

1. Belgium: IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)
2. Croatia: Croatian National Public Health Information System (NAJS)

3. Denmark: Danish Data Health Registries (DK-DHR)
4. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
5. Germany: InGef Research Database (InGef RDB)
6. Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)
7. The Netherlands: Integrated Primary Care Information (IPCI)
8. Norway: Norwegian Linked Health Registry data (NLHR)
9. Spain: The Information System for Research on Primary Care (SIDIAP)
10. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
11. 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 the proposed research question while covering different regions of Europe.

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

In terms of relevance, the selection of data sources was based on the availability of data on the selected GLP-1 RA ingredients to perform the described analyses. In addition, the data sources were chosen considering their ability to support timely Institutional Review Board (IRB) approvals, thus ensuring alignment with the timeline established by stakeholders for the conduct of this study.

The DARWIN EU® portal, as well as information from the onboarding documents, was used to assess whether data sources had information on GLP-1 RAs. Data within the DARWIN EU® portal is kept up to date by extracting the release dates for each dataset in the network and monitoring when data are out-of-date with the expected refresh cycle (typically quarterly or half-yearly). In addition, it is important to have a clear understanding of the time covered by each released data source, as this can vary across different domains. To facilitate this, the C*DMOnboarding* (and *Achilles*) packages (8) contain a 'data density' plot. This plot displays the number of records per OMOP domain monthly. This allows for the obtention of insights relating to when data collection started, when new sources of data were added, and until when data was included. In addition, at the time of inviting data partners, they were informed about the study objectives and asked whether they could participate in the study.

More general-purpose diagnostic tools, *CohortDiagnostics* (9) and *DrugExposureDiagnostics* (10), have been developed. The *CohortDiagnostics* package provides additional insights into cohort characteristics, record counts, and index event misclassification. The *DrugExposureDiagnostics* package evaluates ingredient-specific attributes and patterns in drug exposure records. Upon finalisation of the study protocol and creation of the disease and drug cohorts of interest by DARWIN EU® Coordination Centre, these packages will be executed in each data source by each data partner.

Data source justification and key characteristics

Initiation of GLP-1 RA medicines may be given by a specialist. However, GLP-1 RA medicines are often prescribed by primary care practitioners. In line with type 2 diabetes mellitus patients, users of GLP-1 RA are often followed-up and controlled in primary care settings (referring to GPs, nurses, pharmacists, etc., but not specialists). With this in mind, we selected the following data sources, which cover the healthcare settings of interest:

CPRD GOLD, DK-DHR, IPCI, IQVIA DA Germany, IQVIA LPD Belgium, and SIDIAP were invited to participate in study P4-C2-013.

FinOMOP-THL, HI-SPEED, InGef RDB, NAJS, and NLHR were invited to participate in study P4-C2-014.

Objectives 5 and 6 will be carried out using the previously mentioned data sources, subject to feasibility of execution.

CPRD GOLD, IPCI, IQVIA LPD Belgium, IQVIA DA Germany, and SIDIAP were included in the original P3-C1-008 study. These data sources captured the relevant data necessary to answer the research questions of this study, cover the healthcare settings where the drugs of interest may be used (i.e., primary care), and they have recent data covering the 10-year study period.

In addition, DK-DHR, FinOMOP-THL, HI-SPEED, NAJS, InGef RDB, and NLHR have been added, as they also covered the required setting for the study and have an adequate estimated number of people with GLP-1 RA records. All data sources, except HI-SPEED and NLHR, have recent data covering the 10-year study period; however, HI-SPEED and NLHR were considered of interest because of the nationwide geographical areas they represent (Sweden and Norway, respectively).

IQVIA DA Germany and InGef RDB data sources have an overlap of patients. While the number of active subjects is larger in InGef RDB, this data source has the limitations of 1) only recording outpatient diagnoses per quarter of the year, 2) no complete information on route of administration/dose and duration, and 3) no dose calculation for drug forms besides oral tablets can be done. The last few months before the date of data extraction from IQVIA DA Germany present an artefactual decrease in the denominator which leads to overestimation of the incidence and prevalence estimates. Both data sources have been included for this study.

The selected data sources fulfil the criteria required for a population-level drug utilisation study, as all these data sources have counts for most of the GLP1-RA treatments in the order of thousands of users and a consistent denominator (except for IQVIA LPD Belgium and IQVIA DA Germany), allowing for accurate incidence estimation, and quick approvals/governance, enabling rapid analyses. (**Table 1**)

Information on the data sources planned for use in this study is provided in **Annex II**.

Table 1. Description of the selected data sources.

Country	Name of Data source*	Justification for Inclusion	Type of Data	Number of active subjects	Data lock point for the last update at the protocol stage**
Belgium	IQVIA LPD Belgium	All data sources are population-level (national or regional coverage) that include primary care/outpatient specialist settings where GLP-1 RA	Outpatient General Practitioner (GP) Care	188800	31/03/2025
Croatia	NAJS		Registry	4303700	08/02/2025
Denmark	DK-DHR		Registry	5984000	18/01/2025
Finland	FinOMOP-THL		Registry	5696800	01/10/2024
Germany	InGef RDB		Claims	7665200	01/04/2025

Country	Name of Data source*	Justification for Inclusion	Type of Data	Number of active subjects	Data lock point for the last update at the protocol stage**
	IQVIA DA Germany	are expected to be prescribed.	Outpatient GP Care	4484200	31/12/2024
Norway	NLHR		Registry	6945300	29/10/2024
The Netherlands	IPCI		Outpatient GP Care	1332500	15/04/2025
Spain	SIDIAP		Outpatient GP Care	5951200	31/12/2024
Sweden	HI-SPEED		Registry, Outpatient GP Care, Inpatient Care	10563700	10/09/2024
The United Kingdom	CPRD GOLD		Outpatient GP Care	2632700	01/07/2025

*Data sources: IQVIA LPD=IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium); NAJS=Croatian Institute of Public Health; DK-DHR=Danish Data Health Registries; FinOMOP-THL=Finnish Care Register for Health Care; InGef RDB= InGef - Institute for Applied Health Research Berlin GmbH Research Database; IQVIA DA=IQVIA Disease Analyzer Germany (IQVIA DA Germany); NLHR= Norwegian Linked Health Registry data; IPCI= Integrated Primary Care Information; SIDIAP=The Information System for Research in Primary Care; HI-SPEED= Health Impact - Swedish Population Evidence Enabling Data-linkage; CPRD GOLD=Clinical Practice Research Datalink GOLD.
** Data locks will be updated to the most recent available data at the time of the study execution. GP, general practitioner.

9.3. Study period

The study period will be 01/01/2015 up until the latest data available for each data partner. **Table 1** shows the data lock points for the latest updates of all data sources.

It should be noted that in NLHR and HI-SPEED, the availability of the data starts from 2018 and ends respectively in October and September 2024.

9.4. Follow-up

To determine the incidence and prevalence estimates (Objective 1), we require the appropriate population and their contributed observation time to first be identified (i.e., the denominator population):

- For prevalence estimates, study participants in the denominator population will begin contributing to the denominator on the respective date of the latest of the following: 1) study start date or 2) date at which the observation period starts. Participants will stop contributing person time at the earliest date of the following: 1) end of available data in each of the data sources or 2) date at which the observation period of the specific person ends.

  An example of entry and exit into the denominator population is shown in **Figure 2**. In this example, person ID 1 and 3 are included as denominators at the study start date as both are being observed in the database from a prior date. Person ID 2 enters the study at the date of starting their period of observation, which was later than the study start date. Person ID 1 and 2 will be followed until the study end date (end of available data in each of the data sources) whilst Person ID 3 leaves when exiting the database (the end of observation period). Lastly, person ID 4 has two observation periods in the database. The first period contributes time from study start until end of observation period, the second starts contributing time again on the date of their second observation period start and exits at study end date.

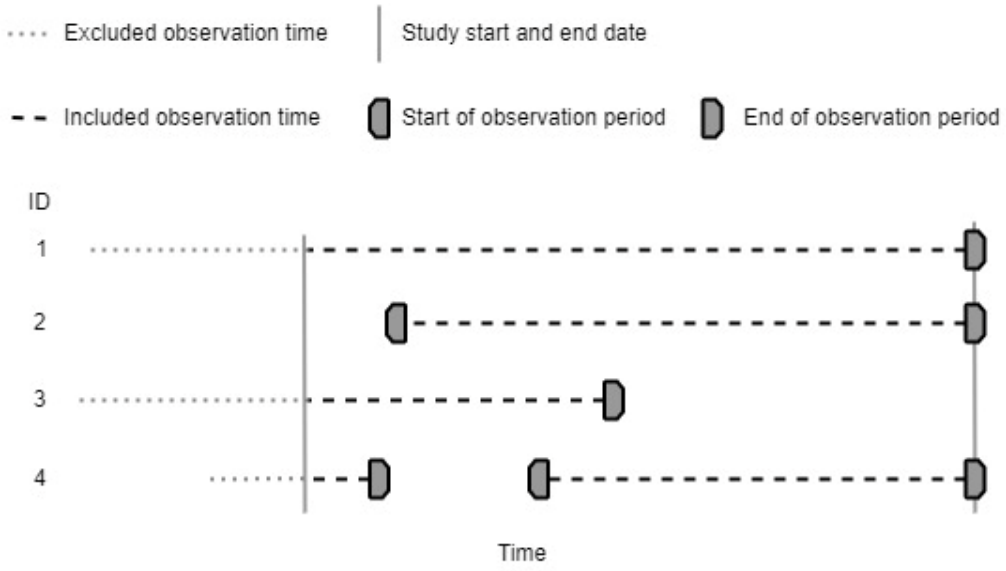


Figure 2. Included observation time for the denominator population in prevalence estimates.

- For incidence rates, an additional criterion is required to ensure that we are capturing an incident use (i.e., new episode of drug use, with a washout window of the specific GLP-1 RA of interest [-365, -1 days]). Thus, follow-up will start from the date they have reached at least 365 days of data availability. Thus, the denominator population will begin contributing person time on the respective date of the latest of the following: 1) study start date, 2) date at which the observation period has already sufficient prior history. Participants will stop contributing person-time at the earliest date of the following: 1) end of available data in each of the data sources or 2) date at which the observation period of the specific person ends. Additionally, when incidence rates are stratified by indication cohort or prescriber speciality, individuals in the denominator will stop contributing person-time from the date of their first prescription for the medicine of interest until one year after their last prescription of that episode. In other words, if a patient with an indication for type 2 diabetes mellitus starts dulaglutide, they will no longer contribute to the denominator of other indication cohorts for dulaglutide from that date until one year after they stop using dulaglutide, as they will no longer be eligible to contribute to the numerator.

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

For the characterisation of new GLP-1 RA users (Objective 2), follow-up will start on the date of the first GLP-1 RA prescription/dispensation (with no prior use of the respective GLP-1 RA drug in the prior 365 days) and will be followed until the end of treatment episode, loss to follow-up, death, or end of observation period (the latest available data).

For the characterisation of GLP-1 RA utilisation pattern switching of substances (Objectives 3 and 5), follow-up will start on the date of the first prescription/dispensation of a GLP-1 RA ingredient (with no prior use of any other GLP-1 RA drug in the prior 365 days) and will be followed until the earliest of loss to follow-up, death, or end of observation period (the latest available data), after switching to other ATC class level drugs described in **Table 4**, or discontinuation of the GLP-1 RA treatment not followed by other ATC class level drug, whatever occurs first. A visual representation is available in **Figure S6**.

For the characterisation of the GLP-1 RA utilisation pattern within-substance switching of strength (Objectives 4 and 6), follow-up will start on the date of the first prescription/dispensation of dulaglutide, semaglutide, liraglutide, tirzepatide, or the brand of interest defined in **Table 5** (with no prior use of their respective GLP-1 RA ingredient in the prior 365 days), and will be followed until the earliest of loss to follow-up, death, end of observation period (the latest available data), or discontinuation of use of the specific GLP-1 RA ingredient, whichever occurs first.

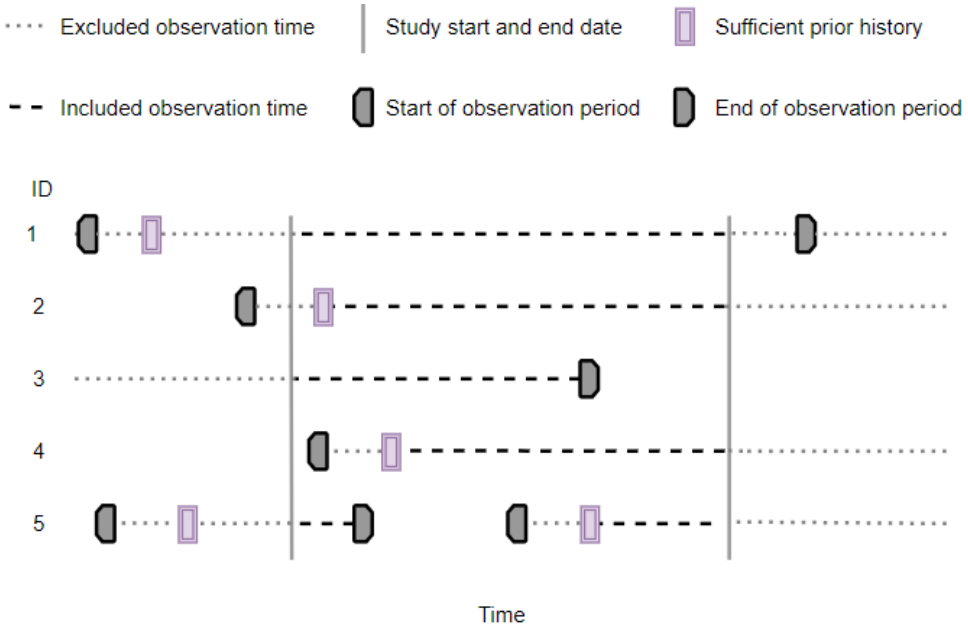


Figure 3. Included observation time for the denominator population in incidence rates.

9.5. Study population with inclusion and exclusion criteria

Individuals with an invalid age (i.e., either no record of birth year and/or an age ≥150 years old) and/or invalid sex (i.e., missing sex) will be excluded from the analysis. Specific inclusion and exclusion criteria for each of the study objectives are described below:

Objective 1: Incidence and prevalence of prescriptions of GLP-1 RA

The study cohort will comprise all individuals present in the database during the study period (i.e., 2015 to the most recent data lock reported).

Additional eligibility criteria for calculation of incidence rates will be applied, where a minimum of 365 days of data availability and no prior use of the respective GLP-1 RA during the previous 365 days will be required.

Objective 2: Characterisation of new GLP-1 RA users

All individuals with a first prescription/dispensation of a GLP-1 RA during the study period, with a minimum of 365 days of data availability and no prior use of the respective GLP-1 RA during the previous 365 days at the time of the treatment start.

Objective 3: Characterisation of GLP-1 RA utilisation pattern for GLP-1 RA switching of substances [overall period]

All individuals with a first prescription/dispensation of a GLP-1 RA during the study period, with a minimum of 365 days of data availability and no prior use of any type of GLP-1 RA during the previous 365 days at the time of GLP-1 RA treatment.

Objective 4: Characterisation of GLP-1 RA utilisation pattern for GLP-1 RA within-substance switching of strength [overall period]

All individuals with a first prescription/dispensation of liraglutide, dulaglutide, semaglutide, or tirzepatide during the study period, with a minimum of 365 days of data availability and no prior use of the respective GLP-1 RA during the previous 365 days at the time of the treatment start.

Objective 5: Characterisation of GLP-1 RA utilisation pattern for GLP-1 RA switching of substances between 2015–2020, and annually between 2021 and 2025

All individuals with a first prescription/dispensation of a GLP-1 RA during the study period, with a minimum of 365 days of data availability and no prior use of any type of GLP-1 RA during the previous 365 days at the time of GLP-1 RA treatment.

Objective 6: Characterisation of GLP-1 RA utilisation pattern for GLP-1 RA within-substance switching of strength between 2015–2020, and annually between 2021 and 2025

All individuals with a first prescription/dispensation of liraglutide, dulaglutide, semaglutide, or tirzepatide during the study period, with a minimum of 365 days of data availability and no prior use of the respective GLP-1 RA during the previous 365 days at the time of the treatment start.

9.6. Variables

9.6.1 Exposure

For this study, the exposure of interest is use of GLP-1 RA during the study period (**Table 2**).

Table 2. GLP-1 RA ingredients included.

Ingredient name*	ATC	Concept ID
Exenatide	A10BJ01	1583722
Liraglutide	A10BJ02	40170911
Lixisenatide	A10BJ03	44506754
Dulaglutide	A10BJ05	45774435
Semaglutide	A10BJ06	793143
Tirzepatide	A10BX16	779705

*This study will not include the GLP-1 RA ingredients albiglutide (it is no longer authorised in the EU and counts are ≤1,000 users/data source) and beinaglutide (it has not been approved for use in the EU).
ATC, Anatomical Therapeutic Chemical.

Records containing a GLP-1 RA ingredient in combination with additional active ingredient(s) (e.g., liraglutide with insulin) will be permitted/included in the study.

Objective 1:

GLP-1 RA will be assessed at the class level (i.e., overall GLP1- RA use), by ingredient (**Table 2**), and by brand (**Table 3**). Brands include Wegovy, Ozempic, Victoza, Saxenda, Rybelsus, Trulicity, and Mounjaro, or, alternatively, a cohort indicating that brand was different than the previously defined or not recorded.

Visual representations are available in **Figure S1** – **Figure S4**.

Table 3. Brand cohorts for incidence and prevalence analysis.

Ingredient name	Cohorts by brand
Semaglutide	Wegovy
	Ozempic
	Rybelsus

Ingredient name	Cohorts by brand
	No brand recorded for semaglutide.
Liraglutide	Victoza
	Saxenda
	Other brands/no brand recorded for liraglutide.
Dulaglutide	Trulicity
	No brand recorded for dulaglutide.
Tirzepatide	Mounjaro
	No brand recorded for tirzepatide.

Objective 2:

Characterisation of new users will be done at the class level (i.e., overall GLP1- RA use), by ingredient (**Table 2**), indication cohorts and by brands as described in **Table 3**.

A visual representation in available in **Figure S5**.

Objective 3 and 5:

GLP-1 RA switching of substances: switching will be described starting from the first prescription of a GLP-1 RA ingredient with no prior use of any other GLP-1 RA ingredient in the prior 365 days (ingredients included in **Table 2**), to one of the following pre-defined cohorts.

- Use of other GLP-1 RA ingredients (i.e., exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide, or tirzepatide cohorts).
- Use of ATC3 and ATC4 class level cohorts of antidiabetic medicines included in **Table 4** (i.e., A10BA, A10BB, A10BD, A10BF, A10BG, A10BH, A10BK, Other A10B, A10A, and A08A cohorts).
- No further medication: defined as no more medication from GLP-1 RA class nor antidiabetic ATC classes described in **Table 4**).

Visual representations are available in **Figure S6** and **Figure S7**.

Table 4. Other second-line antidiabetics at ATC3 or ATC4 level class and insulin at ATC3 level class for the characterisation of GLP-1 RA switching of substances.

ATC drug class level	Name
A10BA	Biguanides
A10BB	Sulfonylureas
A10BD	Combinations of oral blood glucose lowering drugs
A10BF	Alpha glucosidase inhibitors
A10BG	Thiazolidinediones
A10BH	DPP-4 inhibitors
A10BK	SGLT2 inhibitors
Other A10B	The remaining drugs in A10B class (i.e., excluding GLP-1 RA ingredients and excluding ingredients from the above groups).
A10A	Insulin drugs
A08A	Antiobesity preparations, excluding diet products. A08AA62 (bupropion and naltrexone) and A08AA56 (ephedrine, combinations) will be excluded from A08A class.

ATC, Anatomical Therapeutic Chemical; DPP-4, dipeptidyl peptidase 4 inhibitors; GLP-1 RA, Glucagon-like peptide-1 receptor agonists; SGLT2, sodium-glucose co-transporter 2 inhibitors.

Objective 4 and 6:

GLP-1RA within-substance switching of strength: The analysis of switching of strength will be restricted to within the same ingredient. The strength-level cohorts will be done by ingredient-strength or by ingredient-strength-brand (when brand is available) cohorts described in **Table 5** (limited to semaglutide, dulaglutide, liraglutide, and tirzepatide). Brands will be limited to Wegovy, Ozempic, Victoza, Saxenda, Rybelsus, Trulicity, and Mounjaro. Other brands will be included in the 'Other brands/no brand recorded' cohort. A visual representation in available in **Figure S8A**. These analyses will be calculated using incident user (new users of the medicine of interest [e.g., individuals with no use of dulaglutide in the prior 365 days when estimating incidence of dulaglutide use].

Within the same ingredient, we will see the switching of a specific strength (with brand information when available) 1) to the same brand with another strength, 2) to another brand with strength information, 3) to another strength with an unknown brand, or 4) to stop of GLP-1 treatment (see a visual representation in **Figure S8B** and in **Figure S9**).

Table 5 describes the pre-defined strength-levels within each specific ingredient/brand.

Table 5. Pre-defined cohorts of strength-levels.

Type of pre-defined strength-level cohort	Cohort name	Strength switch
Semaglutide Brand	Wegovy (semaglutide [injection])	- 0.25 mg, 0.5mg, 1 mg, 1.7 mg, ≥2.4 mg
Semaglutide Brand	Ozempic (semaglutide [injection])	- 0.25 mg, 0.5 mg, 1 mg, ≥2 mg
Semaglutide Brand	Rybelsus (semaglutide [oral]])	- 1.5 mg, 3 mg, 4 mg, 7 mg, 9 mg, ≥14 mg
Semaglutide Brand	No brand recorded for semaglutide.	- Injection: 0.25 mg, 0.5 mg, 1 mg, ≥2mg - Oral: 3 mg, 5 mg, 7mg, ≥14 mg
Liraglutide Brand	Victoza (liraglutide)	- 0.6 mg, 1.2 mg, ≥1.8 mg
Liraglutide Brand	Saxenda (liraglutide)	- 0.6 mg, 1.2 mg, 1.8 mg, 2.4 mg, ≥3 mg
Liraglutide Brand	Other brands/no brand recorded for liraglutide.	- 0.6 mg, 1.2 mg, 1.8 mg, 2.4 mg, ≥3 mg
Dulaglutide Brand	Trulicity (dulaglutide)	- 0.75 mg, 1.5 mg, 3 mg, ≥4.5 mg
Dulaglutide Brand	No brand recorded for dulaglutide.	- 0.75 mg, 1.5 mg, 3 mg, ≥4.5 mg
Tirzepatide Brand	Mounjaro (tirzepatide)	- 2.5 mg, 5 mg, 7.5 mg, 10 mg, 12.5 mg, ≥15 mg
Tirzepatide Brand	No brand recorded for tirzepatide.	- 2.5 mg, 5 mg, 7.5 mg, 10 mg, 12.5 mg, ≥15 mg

Current tables show the strength based on guidelines. Individual cohorts will be created for each strength within the specific ingredient/brand based on the individual products prescribed in the individual data sources.

9.6.2 Outcome

NA.

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

Age:

Age at index date will be calculated using January 1st of the year of birth as proxy for the actual date of birth.

Covariates for stratification in Objective 1 (Incidence/prevalence):

- Age groups [years]: <18, 18–44,45–54, 55–64, 65+, with assessment window at index date: [0, 0]
- Sex: male/female, with assessment window at index date: [0, 0]
- Indication cohorts [only for incidence]:

  Indication for type 2 diabetes mellitus (T2DM) and obesity will be defined as following:

  - T2DM, defined as a diagnosis of T2DM before new use of GLP-1 RA, with assessment window at index date: [-Inf, 0]. In DK-DHR, indications are registered at index date, therefore, their indications will need to be evaluated at index date [0,0].
  - Obesity, defined as diagnosis [based on presence of clinical codes indicating obesity] or measurement, any time during the observation period. Measurements that qualified as obesity were body mass index (BMI) values between 30 and 60 kg/m2 or, as an alternative when height was not available, weights values between 120 to 200 kg or 265 and 440 pounds, with assessment window at index date: [-Inf, Inf]. In DK-DHR, indications are registered at index date, therefore, their indications will need to be evaluated at index date [0,0].

  Incidence calculations of each of the GLP-1 RA will be also stratified by the combination of type 2 diabetes mellitus (T2DM) and obesity, creating the following four cohorts:

  1. Users of GLP-1 RA with diagnosis of T2DM and obesity
  2. Users of GLP-1 RA with diagnosis of T2DM and excluding obesity
  3. Users of GLP-1 RA with diagnosis of obesity and excluding T2DM
  4. Users of GLP-1 RA excluding obesity and T2DM
- Prescriber speciality [where available and only for incidence]: Prescriber speciality, such as family medicine, endocrinologist, paediatrics, etc., will be used to stratify incident users of each of the GLP-1 RA medicines. The assessment window will be at index date: [0, 0]
- Calendar month

Covariates for stratification in Objective 2 (Characterisation of new GLP-1 RA users)

- Overall study period, and calendar year.
- Indication cohorts (T2DM/obesity) as described above.

Covariates described in Objective 2 (Characterisation of new GLP-1 RA users)

- Age [years] (median, interquartile range [IQR]) and stratified by age groups (number of subjects (%): <18, 18–44,45–54, 55–64, 65+). Assessment window at index date: [0, 0]
- Sex: male/female, with assessment window at index date: [0, 0]
- Body mass index (BMI) [kg/m2] (median [Q25 - Q75], only among those with a BMI record) and stratified by BMI groups (number of subjects (%): ≤26, 27 – 29, 30 – 34, 35 – 39, ≥40 kg/m2 or None [i.e., percentage of individuals with missing BMI]). We will use the most recent record of BMI within the last five years respectively from index date. Values below 10 and over 60 kg/m2 were truncated to diminish the risk of outliers. Assessment window will be between 5 years before (and at) index date: [-5 years, 0]

- Comorbidities (n, %): including hypertension, myocardial infarction[1], stroke, MACE (major adverse cardiovascular event: composed by myocardial infarction and stroke), heart failure[1], atrial fibrillation, peripheral arterial disease, peripheral vascular disease[1] , cerebrovascular disease[1], cardiovascular disease (CVD), chronic kidney disease and renal impairment[1], T2DM, dyslipidaemia (composed of hypercholesterolaemia and hypertriglyceridaemia), thyroid disorders, gallbladder disorders, Alzheimer's disease, neuropsychiatric diseases (i.e., schizophrenia, Parkinson's disease, dementia[1], depression, anxiety, attention-deficit/hyperactivity disorder [ADHD]), asthma, chronic obstructive pulmonary disease (COPD) [1], osteoarthritis, latent autoimmune diabetes in adults (LADA), autoimmune diseases (i.e., rheumatoid arthritis [1], lupus erythematosus[1], psoriasis, psoriatic arthritis, Crohn's disease), peptic ulcer disease[1], liver disease[1], metabolic dysfunction-associated steatotic liver (MASLD), hemiplegia[1], paraplegia[1], sleep disorder, sleep apnoea, polycystic ovary syndrome (PCOS), eating disorders (composed of anorexia and bulimia), cancer (any excluding non-melanoma skin cancer)[1], pancreatic cancer, human immunodeficiency virus (HIV)[1]. Additionally, when available in at least one data source, we will include the individual conditions from the frailty index described in **Table S2** that were not mentioned above (i.e., activity limitation, chronic respiratory disease, falls, hearing impairment, housebound, mobility and transfer problems, requirement for care, sleep disturbance, social vulnerability, urinary incontinence, visual impairment, anaemia and haematinic deficiency, arthritis, dizziness, dyspnoea, foot problem, fragility fracture, heart valve disease, hypotension/syncope, ischaemic heart disease, memory and cognitive problems, osteoporosis, skin ulcer, and urinary system disease).

  [1]Conditions in Charlson comorbidity score.

  For comorbidities with no phenotype available from previous DARWIN EU® studies, the condition will be defined through the use of high-level concepts and their respective descendants.
  The assessment window for comorbidities will be any time before index date: [-Inf, 0].
- Related co-medications (n, %): use of other second-line antidiabetics and insulin at ATC level class group (as described in **Table 4**). Patients are allowed to contribute to more than one ATC level class group. Assessment window will be between 1 year before index date: [-365, -1].
- Initial dose (milligram), assessment window at index date: [0, 0]
- Cumulative dose (milligram) for first continuous treatment episode. Assessment window at index date: [0, end first episode]
- Treatment duration of the first continuous treatment episode: Treatment duration will be calculated as the duration of the first continuous exposure episode, with less than a 30-day gap between prescriptions. Estimations of treatment duration will be summarized providing the median [IQR] treatment duration. For data sources where duration cannot be calculated, due to e.g., missing information on quantity or dosing, treatment duration will not be provided. Assessment window at index date: [0, Inf]
- Proportion of first prescription records mapped to ingredient level in GLP-1 RA users. Assessment window at index date: [0, 0]
- Additionally, we will evaluate the feasibility of including the following covariates during the execution of the diagnostic tools: ethnicity, geographic stratification (e.g., region), and post-partum for women during the last 12 months before first GLP-1 RA use.

<u>Covariates for stratification in Objectives 3 and 4 (Switching)</u>

Analyses will be run over the overall study period.

<u>Covariates for stratification in Objectives 5 and 6 (Switching)</u>

Analyses will be run in the following time-period stratifications:

- Between 2015–2020
- Annually for the years 2021, 2022, 2023, 2024, and 2025.

9.7. Study size

No sample size has been calculated for this study, which will not test a specific hypothesis. The expected number of prescriptions across data sources is expected to be roughly between <1,000 (for lixisenatide in IQVIA LPD Belgium, DK-DHR, FinOMOP-THL, HI-SPEED, InGef RDB, IPCI, and NLHR) and >300,000 (for semaglutide in DK-DHR).

9.8. Analysis

The following analyses will be conducted for Objectives 1 – 6 (**Table 6**).

1) Incidence/prevalence (Population-level DUS): *IncidencePrevalence*.

2) New user characterisation (Patient-level DUS): *CohortCharacterisation* and *DrugUtilisation*.

3–6) Switching (Patient-level characterisation): *Treatmentpatterns* (summarising treatment switching) and *CohortSurvival* (time-to-switch from new GLP-1 RA use to other non-GLP-1 RA treatment, and time-to-switch between strengths).

Table 6. Description of study types and type of analysis.

Study	Study classification	Type of analysis
Population-level DUS	Off-the-shelf	- Population-based incidence rates and prevalence estimates
Patient-level DUS	Off-the-shelf	- Characterisation of patient-level features - Frequency and % of comorbidity/es - Estimation of minimum, p25, median, p75, and maximum initially prescribed or dispensed dose/strength - Estimation of minimum, p25, median, p75, and maximum treatment duration
Patient-level characterisation	Off-the-shelf	- Patient-level treatment patterns - Time-to-event (i.e., time to switch): median time-to-switch survival (based on median survival), average time-on-treatment (based on restricted mean survival), and Kaplan-Meier plots.

DUS, drug utilisation studies.

9.8.1. Federated network analyses

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

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

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

9.8.2. Patient privacy protection

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

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

9.8.3.1. R-packages

This study will be calculated based on OMOP CDM mapped data. We will use the R package *IncidencePrevalence* for the population-level drug utilisation study and the R package *DrugUtilization* for the patient-level drug utilisation study. Additionally, we will use the R packages *TreatmentPatterns* and *CohortSurvival* for the patient-level characterisation.

9.8.3.2. Population-level incidence calculations

New drug users for the different medicines of interest will be captured.

Calendar time for incidence calculations will be based on monthly periods of the index prescription/dispensation in the population level analysis (e.g., 01/01/2015 – 31/01/2015, 01/02/2015 – 30/02/2015, etc.).

Monthly incidence rates of the medicines of interest will be calculated as the number of users per 100,000 person-years of the population at risk of getting exposed during the period for each calendar year. Those study participants who enter the denominator population will then contribute time at risk up to their first new use (prescription or dispensation) during the study period. Alternatively, if they do not have a drug exposure, they will contribute time at risk up as described above in **section 9.4 (Follow-up)**. Incidence rates will be given together with 95% confidence intervals using the exact method.

An illustration of the calculation of incidence use for each of the medicines of interest is shown below in **Figure 4**.

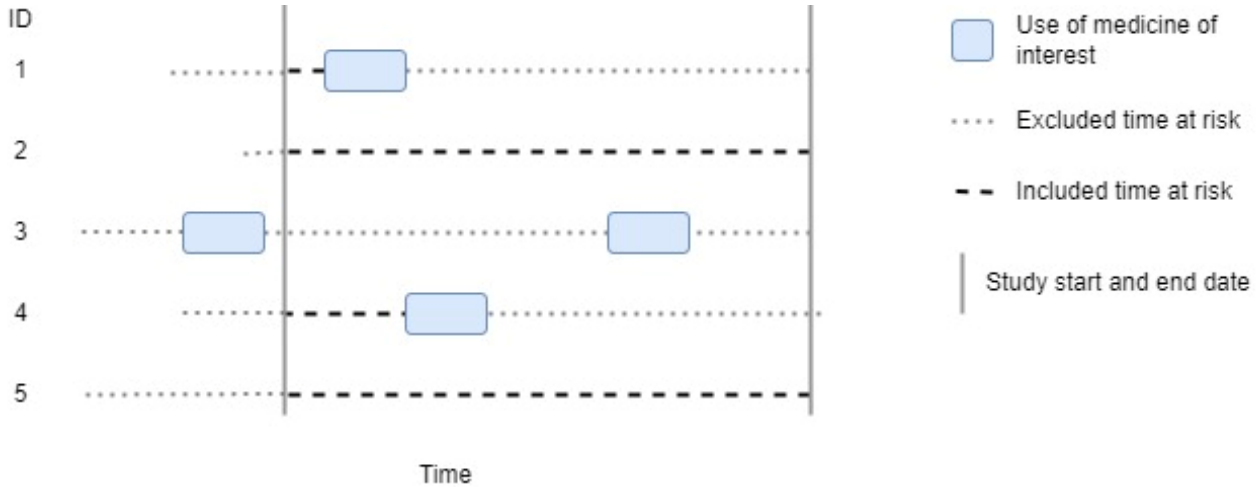


Figure 4. Example for capturing new incidence use of medicines of interest.

Patient ID 1 and 4 contribute time at risk between the study start until when they become incident users of the medicine of interest. Patient ID 2 and 5 contribute time at risk between the study start and end date as no use of medicines of interest is observed between this period nor before the study start date. Patient ID 3 was excluded from the analysis and do not contribute time at risk since a previous use of the same medicine of interest was observed before the study start date.

9.8.3.3. Population-level prevalence calculations

Calendar time for point prevalence calculations will be based on the first day of each calendar month in the population level analysis (e.g., 01/01/2024, 01/02/2024, etc.).

Monthly point prevalence of the medicines of interest will be calculated as the total number of individuals who use the medicine of interest at the start of each calendar month, divided by the population at risk of getting exposed during that point. Therefore, point prevalence gives the proportion of individuals exposed at the first of each calendar month.

Point prevalence will be given together with 95% confidence intervals. Binomial 95% confidence intervals will be calculated using the Wilson Score method for binomial distribution.

An illustration of the calculation of point prevalence use for each of the medicines of interest is shown below in **Figure 5**.

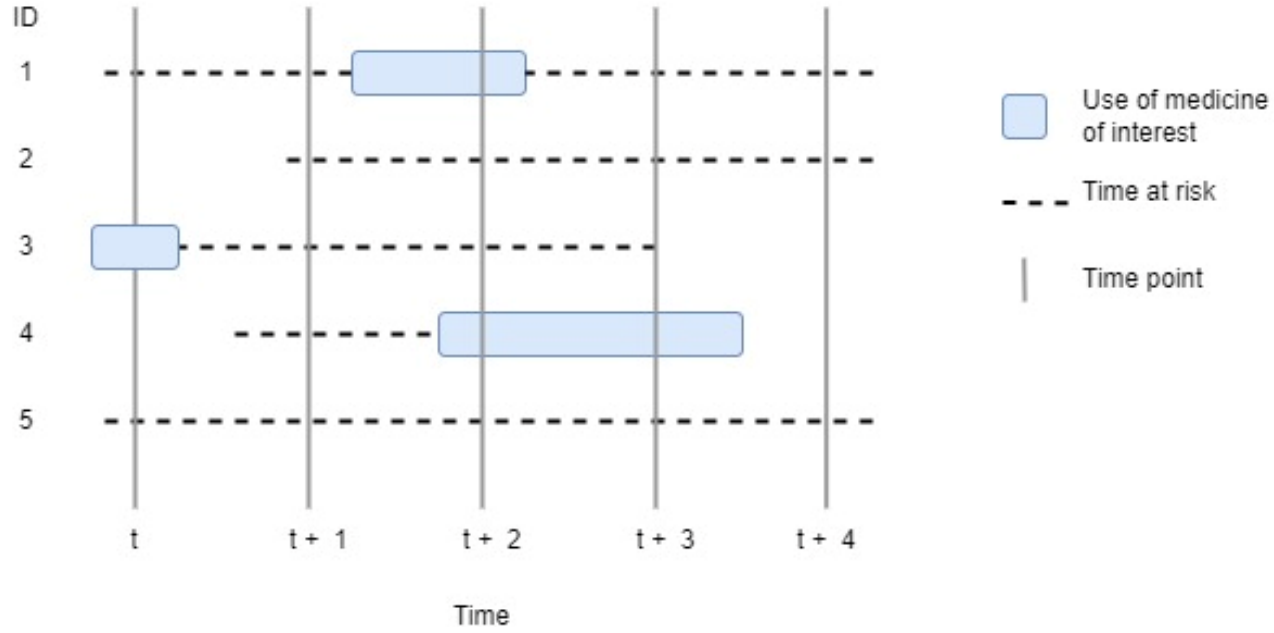


Figure 5. Example for capturing point prevalence use of medicines.

Point prevalence use is captured at the different time points, represented as “t” or “t + n”. In each of the time points, the number of users of the medicine of interest contributes to the numerator. For instance, from the 5 participants, there is only one user for each time point at time “t” and at time “t + 3”, whilst “t + 1” and “t + 4” have no users, and “t + 2” have two users of the medicine of interest. Patient ID 1 and 3 only contribute one time point each: “t” and “t + 2”, respectively. Conversely, Patient ID 4 contributes to two-time points, “t + 2” and “t + 3”, as their use of the medicine of interest continued.

9.8.3.4. Patient-level characteristics on/before index date

New drug users of GLP-1 RA stratified by presence or absence of diagnosis of T2DM and/or diagnosis of obesity will be characterised by age, sex, body mass index, initial dose, cumulative dose, and a list of prespecified indications/comorbidities and related co-medications.

Average age and number of men and women (N, %) at index date will be reported. The list of prespecified indications/comorbidities and related co-medications will be reported using the number of persons and percentage (N, %) with a record within the prespecified time windows. Body mass index will be reported as an average and/or as number of persons and percentage (N, %) among the body mass index categories described in **section 9.6.3**.

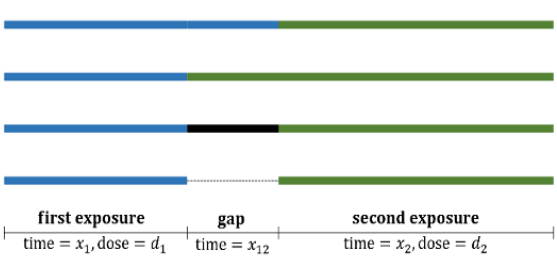
The time widows for covariate assessment are described in **section 9.6.3**.

9.8.3.5. Drug exposure calculations

Drug exposure is the duration of time exposed to a specific ingredient for an individual drug record (i.e., for an individual prescription/dispensation record, how much time that prescription will last). Conversely, drug era is the timeframe in which an individual is persistently considered to be exposed to a specific ingredient. Each drug era can correspond to one or may drug exposures.

For each patient, drug eras will be defined as follows: exposure starts at date of the first prescription after a washout of 90 days. For each prescription, the estimated duration of use is retrieved from the drug exposure table in the CDM, using the start and end date 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 end of the first era and start of the second era is ≤90 days. The time between the two joined eras will be considered as exposed by the first era as shown in **Figure 6**, first row.

Gap era joint mode	**Schematics**	**Dose in between**	**Cumulative dose**	**Cumulative time**
"first"		d_1	$d_1 \cdot (x_1 + x_{12}) + d_2 \cdot x_2$	$x_1 + x_{12} + x_2$
"second"		d_2	$d_1 \cdot x_1 + d_2 \cdot (x_2 + x_{12})$	$x_1 + x_{12} + x_2$
"zero"		0	$d_1 \cdot x_1 + d_2 \cdot x_2$	$x_1 + x_{12} + x_2$
"join"		NA	$d_1 \cdot x_1 + d_2 \cdot x_2$	$x_1 + x_2$

Different drug exposures are differentiated as blue and green. Drug era concatenates different drug exposures based on the gap between them (in our study, two exposures will be concatenated when the gap between them is 90 days or less).

Figure 6. Gap era joint mode.

9.8.3.6. Switching of substances

In objectives 3 and 5, analysis of switching will start at the time of the first new prescription of any GLP-1 RA ingredient with no prior use of any other GLP-1 RA ingredient in the prior 365 days for each person. In objectives 4 and 6, analysis of switching will start at the time of the first new prescriptions of liraglutide, dulaglutide, semaglutide, and tirzepatide.

The treatment pathway will be constructed using the standardised method using the *TreatmentPatterns* R Package. **Figure 7** explains how treatment combination will be defined. The minimum overlap of different treatments to be considered as combination treatment (combinationWindow) is 30 days in the purposed study.

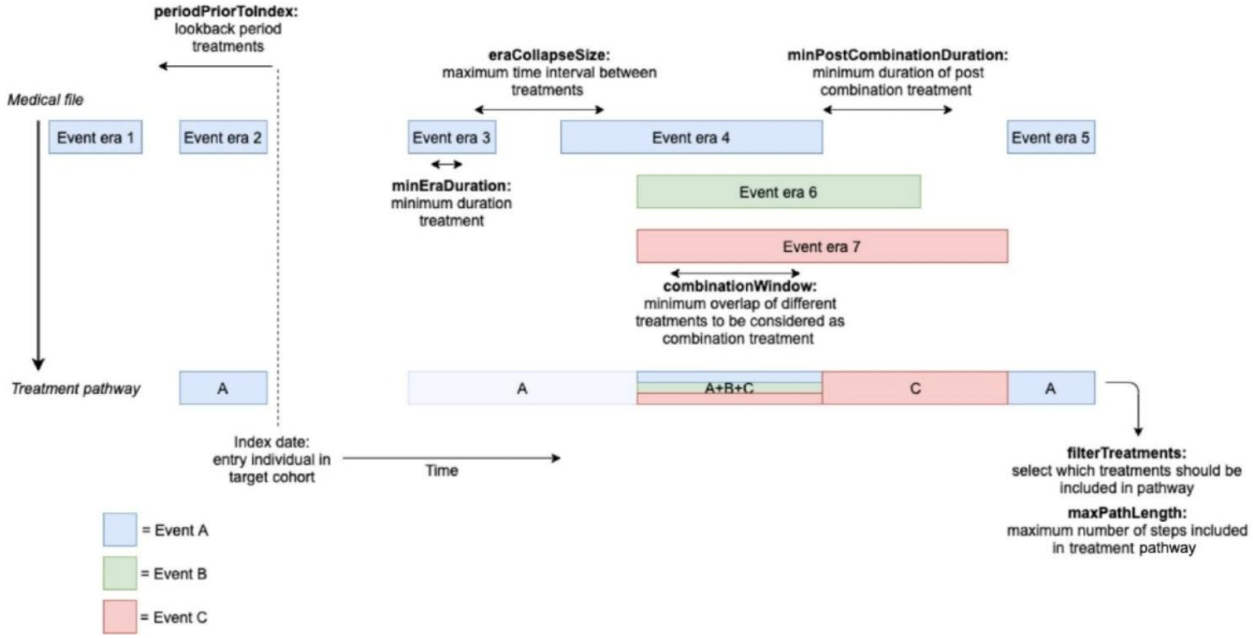


Figure 7. Parameters in *TreatmentPatterns* package

The following parameters will be defined in this study. The target cohort refers to the new user cohorts of GLP-1 RA ingredients (**Table 2**), whereas the event(s) refer to treatment(s) of interest.

We will use a Sankey diagram to illustrate the treatment pathways. To avoid an overwhelming figure, we will only include treatment pathways representing more than 5% people in the entire cohort. Pathways with less than 5% frequency will not be reported.

Table 7. Pathway settings in *TreatmentPatterns*.

Variable in *TreatmentPatterns*	Explanation	Predefined parameter
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 (not relevant as cohort entry start at treatment initiation)
minEraDuration	Minimum time an event era should last to be included in the analysis	1 day
eraCollapseSize	Maximum gap within two eras of the same event cohort which would still allow the eras to be collapsed into one era	0 days (era collapse will be implemented when the cohorts are created)
combinationWindow	Time that two event eras need to overlap to be considered a combination treatment	30 days
minPostCombinationDuration	Minimum time that an event era before or after generated combination treatment should last to be included in the pathway as a separate treatment	30 days
filterTreatments	Select which treatments should be included in pathway first time occurrences of treatments ('First'), remove sequential repeated treatments ('Changes'), all treatments ('All')	First
maxPathLength	Maximum number of treatments included in pathway	10 (will be reduced if number of individuals are low)
Aggregate pathway settings		

Variable in *TreatmentPatterns*	Explanation	Predefined parameter
Individual pathway settings		
groupCombinations	Select to group all non-fixed combinations in one category 'other' in the sunburst plot	TRUE / 10
addNoPaths	Select to include untreated persons without treatment pathway in the sunburst plot	TRUE

9.8.3.7. Time-to-switch

Time-to-switch will be assessed as follows:

- Objectives 3 and 5) New GLP-1 RA use with no prior use of any other GLP-1 RA ingredient in the prior 365 days

Time-to-switch from the new GLP-1 RA use (i.e., from the first episode/drug era) to other medicines will be calculated based on survival analysis and reported using median time-to-switch (based on median survival: time where 50% of the patients in the group had switch to another treatment) and average time-on-treatment (based on restricted mean survival: average time patients are alive and still had not changed the treatment over the study period analysed). Additionally, we will provide Kaplan-Meier curves.

Switching will be censored at the earliest of loss to follow-up, death, end of observation period (the latest available data), after switching to other ATC class level drugs described in **Table 4**, or discontinuation of the GLP-1 RA treatment not followed by other ATC class level drug, whichever occurs first. In the Sankey plot, treatment "STOP" will represent the cessation of the last GLP-1 RA treatment received with no switching to any other of the ATC3 and ATC4 level class cohorts described in **Table 4**.

- Objectives 4 and 6) Switch of strengths within the same ingredient (liraglutide, dulaglutide, semaglutide, or tirzepatide; and with Brand information, when available).

Time-to-switch between strengths (within the same ingredient) will be calculated based on survival analysis and reported using median time-to-switch (based on median survival: time where 50% of the patients in the group had switch to another strength) and average time-on-treatment (based on restricted mean survival: average time patients are alive and still had not changed the strength over the study period analyses). Additionally, we will provide Kaplan-Meier curves.

Switching will be censored at the earliest of loss to follow-up, death, end of observation period (the latest available data), or discontinuation of use of the specific GLP-1 RA ingredient, whichever occurs first. In the Sankey plot, treatment "STOP" will represent cessation of the GLP-1 RA treatment or switch to another GLP-1 RA ingredient.

Differences between restricted mean survival and median survival:

The restricted mean survival is the average event-free survival time over the whole follow-up period (i.e., the average time a patient is alive and free of specific events). In this study, it will be the average time patients are treated with a specific GLP-1 RA medication.

The median survival is the time 50% of the group had the specific event.

The difference between restricted mean survival and median survival is that median survival is the time 50% of the group had the specific event. Thus, there are cases that the median survival result may be null (because it does not reach the 50%), but we will still be able to calculate the restricted mean survival results to understand how many users are still under e.g., "semaglutide treatment" at the end of follow-up.

9.8.4. Output

Output will include the following:

- An Interactive dashboard (shiny app) will be generated by incorporating all the results of the study.
- A PDF report including an executive summary. The final list of tables and figures included in the main report will be agreed upon at a later stage. The list below is an example of the tables and figures that can be included:
    - Table 1. Number of new users of GLP-1 (overall in the last reported year).
    - Table 2. Characteristics of new users of GLP-1 (by GLP-1 RA ingredient and per 11 data sources, and stratified by overall period vs last reported year).
    - Table 3. Initial and cumulative dose (by GLP-1 RA ingredient and per data source and stratified by overall period vs last reported year).
    - Table 4. Time-to-switch from new GLP-1 RA use with no prior use of any other GLP-1 RA ingredient in the prior 365 days.
    - Tables 5-8. Time-to-switch between strengths in liraglutide, dulaglutide, semaglutide, and tirzepatide (independent table by ingredient).
    - Figure 1. Overall incidence rate of GLP-1 RA class in each data source.
    - Figures 2–12. Overall incidence Rates of GLP-1 RA ingredients (independent figure per data source).
    - Figures 13–23. Incidence Rates of GLP-1 RA ingredients by indication cohorts (independent figure per data source).
    - Figures 24. Overall prevalence of GLP-1 RA class in each data source.
    - Figures 25–35. Overall prevalence of GLP-1 RA ingredients (independent figure per data source).
    - Figures 35-101. Within GLP-1 RA switching of substances (6 independent figures per data source).
    - Figures 102–211. Within-substance switching of strength (10 independent figures per data source).

9.9. Missing data

Individuals with a missing sex, year of birth or an invalid age record (e.g., being aged 150 years old) will be excluded from the analysis. Variables used in the study will be based on the recorded diagnoses and prescription codes available in the data, where the lack of a record will be considered as indicating that the patient was not diagnosed/prescribed with the outcome/condition/drug of interest. For data sources where duration cannot be calculated, due to e.g., missing information on quantity or dosing, treatment duration will not be provided. In case of BMI, we will report the percentage of individuals with missing BMI.

9.10. Evidence synthesis

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

10. STRENGTHS AND LIMITATIONS

Our study will have several strengths. It will be based on real-world data with large multinational coverage across Europe, providing a comprehensive overview of GLP-1 RA use in 10 European countries over the past 10-years. Additionally, the analysis will include granular drug utilisation analyses, including brand-level assessments and multiple time windows. The use of the OMOP CDM and development of all analytic code using open-source tools within a federated approach ensures that our approach is fully reproducible and scalable for future efforts. All data sources have previously undergone multiple quality control checks to ensure data quality and to support a clear understanding of the data's characteristics and limitations.

The study will be informed by routinely collected health care data and so data quality issues must be considered. In particular, a recording of a prescription or dispensation does not mean that the patient actually took the drug. The actual reason for prescription of the drug is not recorded in any of the data sources, except for DK-DHR. We will assess indication in the other data sources via a proxy based on pre-defined conditions recorded on or before the date of therapy initiation. Therefore, recording of potential indication may be incomplete. Moreover, prescriber speciality may also not be available in the selected databases. We will include it in the analysis when possible. In addition, assumptions around the duration of drug use will be unavoidable. Calculation of daily and cumulative dose will be based on the formulas developed in the OMOP CDM to match the World Health Organisation Defined Daily Dose (WHO DDD), which were applicable to >85% of drug records of IPCI and CPRD GOLD (among other data sources not included in this study).(11) However, the quality of recording of drug dose might be of varying quality for different data sources. Therefore, data quality checks will be conducted to evaluate the quality of the recording of units, dosage (OMOP drug_exposure tables) for GLP-1 RA in the data sources of the study. Similarly, the documentation of comorbidities, necessary for patient-level characterisation, may vary across data sources.

The data will reflect only publicly reimbursed prescriptions or dispensations. Moreover, marketing status, prescription restrictions, and reimbursement policies might lead to an underestimation of the true incidence/prevalence of GLP-1 RA use due to 1) in data sources based on primary care records (e.g., CPRD GOLD, IPCI, and SIDIAP) that do not contain information on prescription of drugs given by specialists, who may also prescribe GLP-1 RA drugs, or 2) several countries allow for private prescriptions of GLP-1 RA, which are then paid out-of-pocket and are not reimbursed. This might mean that e.g., claims data, or data only capturing prescriptions that are reimbursed by national health systems, are not including those GLP-1 RA records. Although absolute GLP-1 RA use might be underestimated, trends over time will still provide meaningful and actionable information for healthcare professionals and regulators.

Additionally, the results estimated from this study will only reflect the populations from the included data sources. Electronic health records have certain inherent limitations because they were collected for clinical purpose rather than primarily for research use. Consequently, using 11 primary care, outpatient, national registries, and claims data sources from Belgium, Croatia, Denmark, Finland, Germany, the Netherlands, Norway, Spain, Sweden, and the United Kingdom limits generalisability to those countries.

Study specific limitations include:

When applying the requirement of minimum of one year of prior clinical history in incident users, children under one year of age will be excluded by design. As children in this age group are not eligible for GLP-1 RA treatment, this exclusion is not expected to affect the study findings.

Information on brands will be descriptive and exploratory and will only be available for some, but not all, records. The stratification for brand (e.g., Ozempic/Wegovy) might therefore be limited and underestimate the true proportion of the branded products. The extent of this limitation will only be known after the completion of the study.

Database-specific limitations:

In IQVIA LPD Belgium and IQVIA DA Germany, the observation period of the patients is calculated based on the last visit, observation, or interaction of the patient with the health care system. This methodology impacts the individuals considered “at risk” for the different medicines of interest of the study (i.e., the individuals included in the denominator populations) during the latest months of available data from the latest data lock, where healthy and/or non-frequent users of the health care system will not be considered active. Consequently, the denominators that will be used to calculate the incident use of drugs in the population may present an artefactual decrease whilst the incident users will remain, incrementing the incidence and prevalence ratios. To reduce the effect of the artifacts in the denominator, we will remove the last 3 months of data in IQVIA LPD Belgium and IQVIA DA Germany for all the study objectives.

DK-DHR does not have data available on specialty, drug brands, or strength. BMI is largely incomplete, as weight measurements are only recorded in pregnant women. Record of conditions are limited to individuals who had hospital contact. For instance, asthma, depression, anxiety, and other chronic conditions that are mainly treated at the GP will likely be captured in individuals who presented severe cases and were hospitalized. However, drug indications are recorded. Therefore, obesity and T2DM indications can be obtained.

NLHR does not have data available on drug brands or strength. BMI is largely incomplete, as weight measurements are only recorded in pregnant women at the start of pregnancy. Primary care data in NLHR starts from 2008, while drug data and secondary care data start from 2018, and all of datasets are available until the end of 2023. Thus, the study period in NLHR will start in 2018. Likewise, the study period in HI-SPEED starts in 2018. Due to this data availability, incidence rates for NLHR and HI-SPEED will be restricted to 2019–onwards.

In InGef RDB, all outpatient diagnoses are only recorded per quarter of the year, and information on route/dose form and duration is incomplete. Thus, no dose calculation for any dose forms besides tablets can reliably be calculated in InGef RDB. Moreover, exposure days (i.e., days' supply) are not available in InGef RDB and were mapped to 30 days by default, which limits the interpretation of the treatment duration in this database. While InGef RDB includes records of the performance of operations and procedures, no results of procedures or clinical values are captured. Therefore, BMI measurements cannot be captured.

Similarly, NAJS does not contain detailed recordings of dose/duration, as a substantial proportion of records are mapped at ingredient level, which limits its inclusion in the drug utilisation analyses.

While IQVIA DA Germany has geographical/population overlap with InGef RDB, there are also differences in data type/setting. Differences in incidence/prevalence and GLP-1 RA utilisation between these data sources are therefore expected (to some degree) and care should be taken when comparing results from the data sources covering the same country.

11. REFERENCES

1. EMA. Recommendations of the Executive Steering Group on Shortages and Safety of Medicinal Products on shortage of Glucagon-Like Peptide-1 (GLP-1) receptor agonists. EMA/143883/2024 2024 [Available from: https://www.ema.europa.eu/en/documents/other/mssg-recommendations-shortage-glucagon-peptide-1-glp-1-receptor-agonists_en.pdf.

2. DARWIN EU® Coordination-Centre. DARWIN EU® - Drug Utilisation Study on GLP-1 Receptor Agonists. EUPAS ID: EUPAS1000000223 2024 [Available from: https://catalogues.ema.europa.eu/node/4124/resources.

3. EMA. Public information on medicine shortages 2024 [Available from: https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/medicine-shortages-availability-issues/public-information-medicine-shortages#ema-shortages-catalogue-11001.

4. EMA. EU actions to tackle shortages of GLP-1 receptor agonists 2024 [Available from: https://www.ema.europa.eu/en/news/eu-actions-tackle-shortages-glp-1-receptor-agonists.

5. Blind E, Janssen H, Dunder K, de Graeff PA. The European Medicines Agency's approval of new medicines for type 2 diabetes. Diabetes Obes Metab. 2018;20(9):2059-63.

6. EMA. Medicines catalogue in the EU [Available from: https://www.ema.europa.eu/en/medicines.

7. EMA. Mounjaro - opinion on variation to marketing authorisation (tirzepatide) 2023 [updated 10/11/2023. Available from: https://www.ema.europa.eu/en/medicines/human/variation/mounjaro.

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

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

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

11. Burkard T, Lopez-Guell K, Gorbachev A, Bellas L, Jodicke AM, Burn E, et al. Calculating daily dose in the Observational Medical Outcomes Partnership Common Data Model. Pharmacoepidemiol Drug Saf. 2024;33(6):e5809.

12. ANNEXES

ANNEX I. Further details on and visuals of the outcomes of the study

Objective 1 – Point prevalence estimates

1) *To determine the incidence and prevalence of prescriptions of the GLP-1 RA medicines (overall, by ingredient, by pre-specified brand) during the last 10 years of available data, stratified separately by age, sex, indication [only for incidence], and prescriber speciality (where available) [only for incidence], and by calendar month in each of the data sources.*

These analyses will be calculated using prevalent users (i.e., persistent users, as shown in **Figure 1A** from **section 9.1.**).

We will generate point prevalence estimates both for the overall study period and monthly, separately by overall GLP-1 RA use, by ingredient, and by pre-specified brand. These will be then stratified by age and sex, as depicted in **Figure S1**.

The monthly trends of use can show the increased demands of the different GLP-1 RA drugs overall, by ingredient, and by the pre-specified brands. Data may also show decreases that can be consistent with a declared shortage.

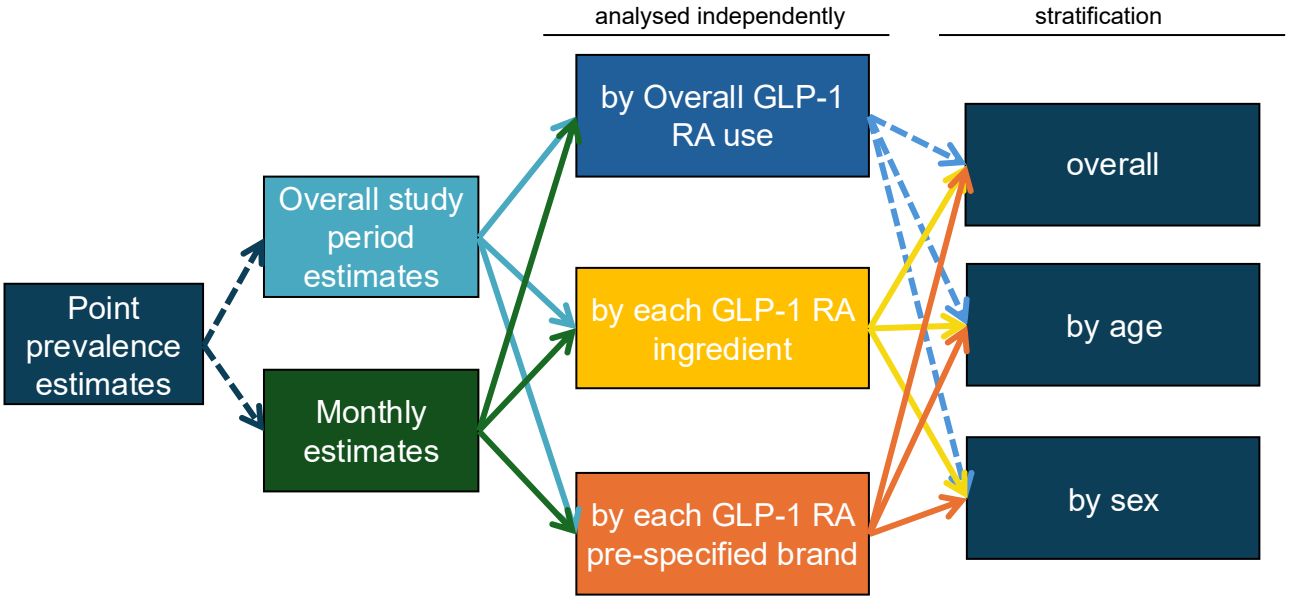
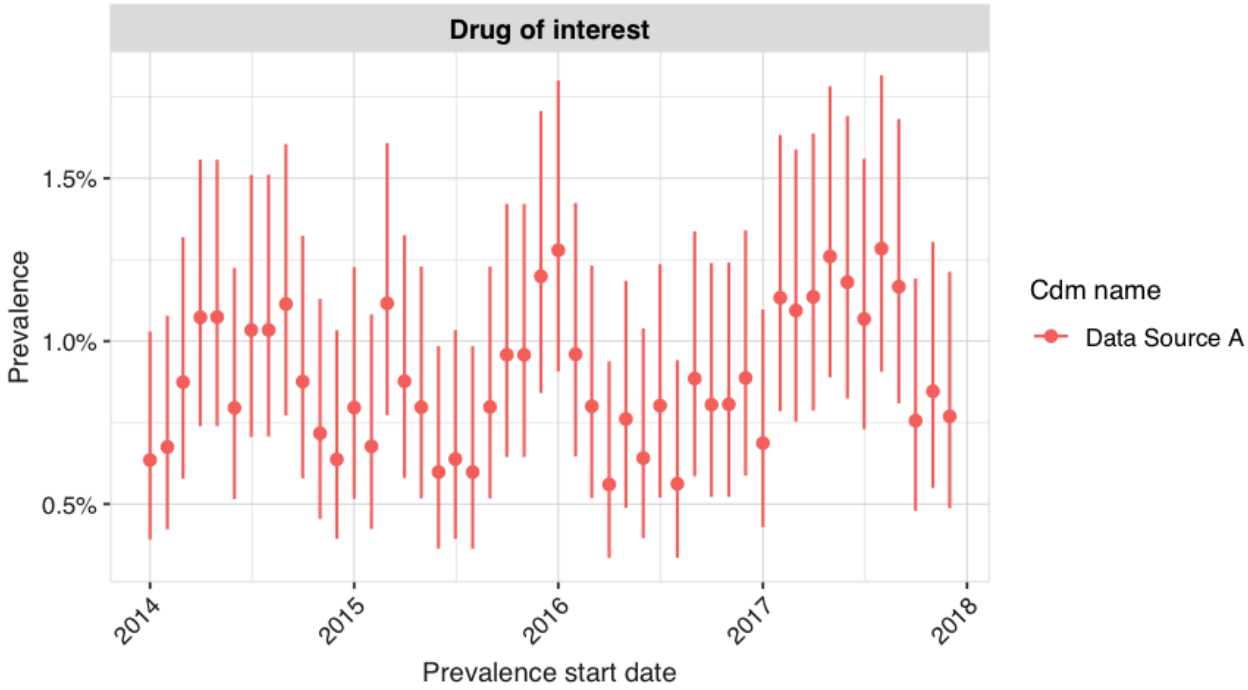


Figure S1. Diagram of the analyses included in Objective 1 – To determine point prevalence estimates.

An example of the monthly point prevalence plots that will be generated is presented in **Figure S2**.

A)



B)

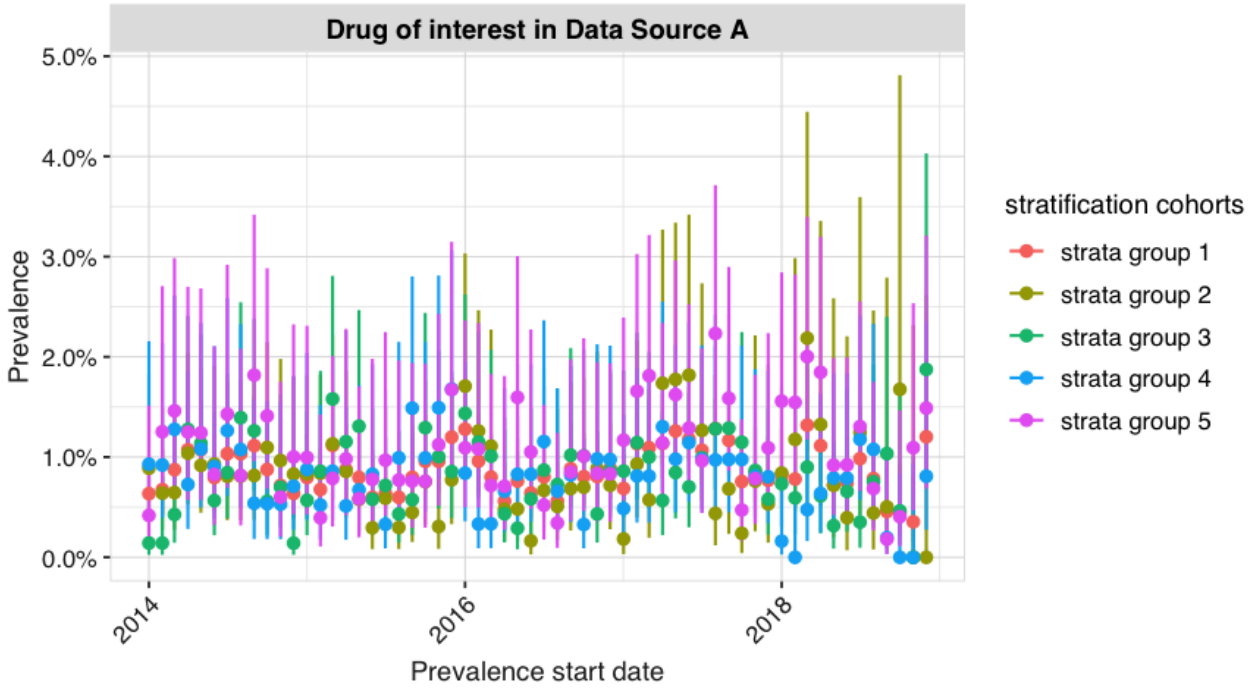


Figure S2. Mock representations of the prevalence plots that will be generated in *Objective 1 – To determine point prevalence estimates*.

Objective 1 – Incidence rates estimates

1) *To determine the incidence and prevalence of prescriptions of the GLP-1 RA medicines (overall, by ingredient, by pre-specified brand) during the last 10 years of available data, stratified separately by age, sex, indication [only for incidence], and prescriber speciality (where available) [only for incidence], and by calendar month in each of the data sources.*

These analyses will be calculated using incident users (new users of the medicine of interest [e.g., individuals with no prior use of dulaglutide when estimating incidence of dulaglutide use], as shown in **Figure 1B** from **section 9.1.**).

We will generate incidence rates for the overall study period and monthly, separately by overall GLP-1 RA use, by ingredient, and by pre-specified brand. These will be then stratified by age, sex, indication, and speciality (when available), as depicted in **Figure S3**.

The monthly trends of use can show the increased demands in new users of the different GLP-1 RA drugs overall, by ingredient, and by the pre-specified brands. Data may also show decreases that can be consistent with a declared shortage.

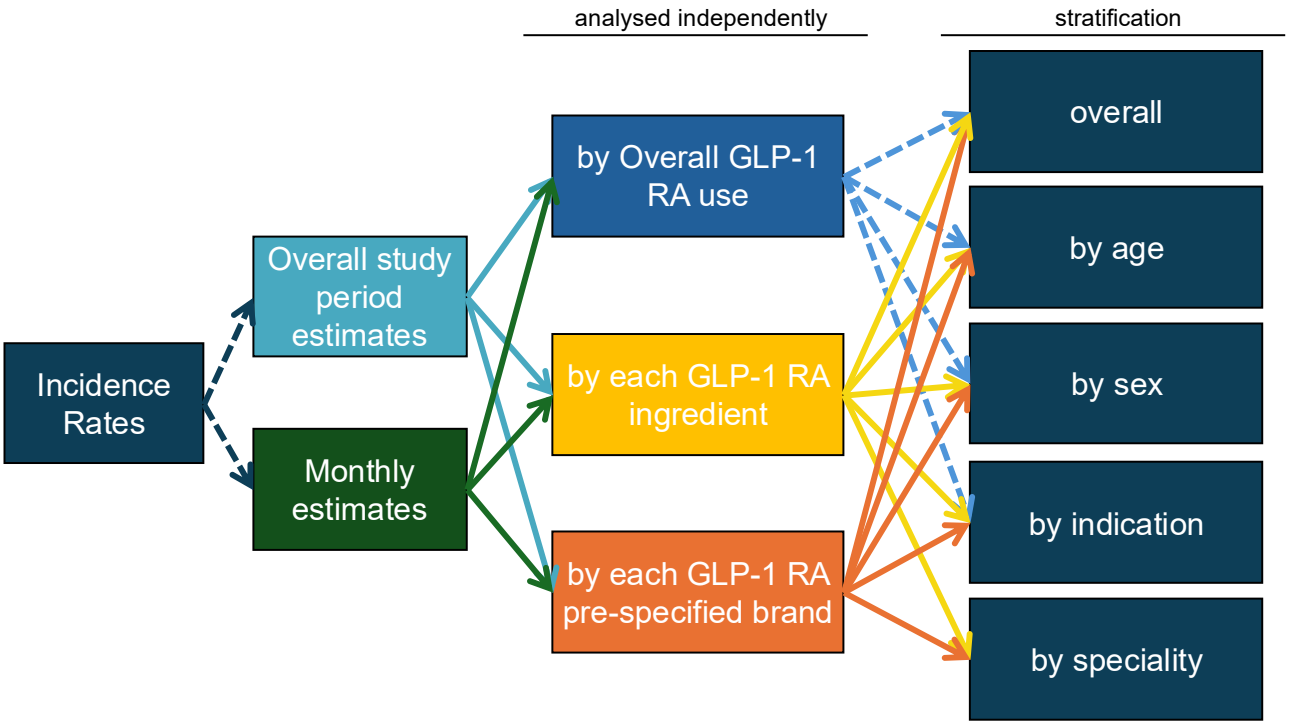
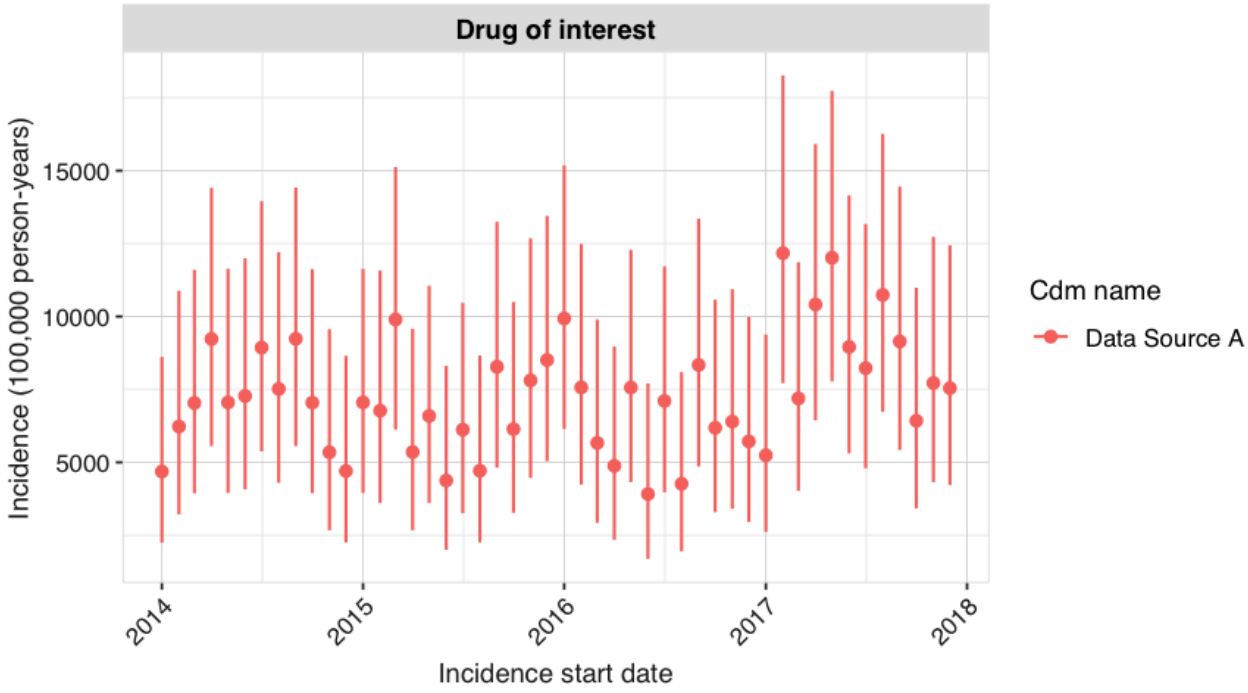


Figure S3. Diagram of the analyses included in *Objective 1 – To determine incidence rates estimates*.

An example of the monthly incidence rates plots that will be generated is presented in **Figure S4**.

A)



B)

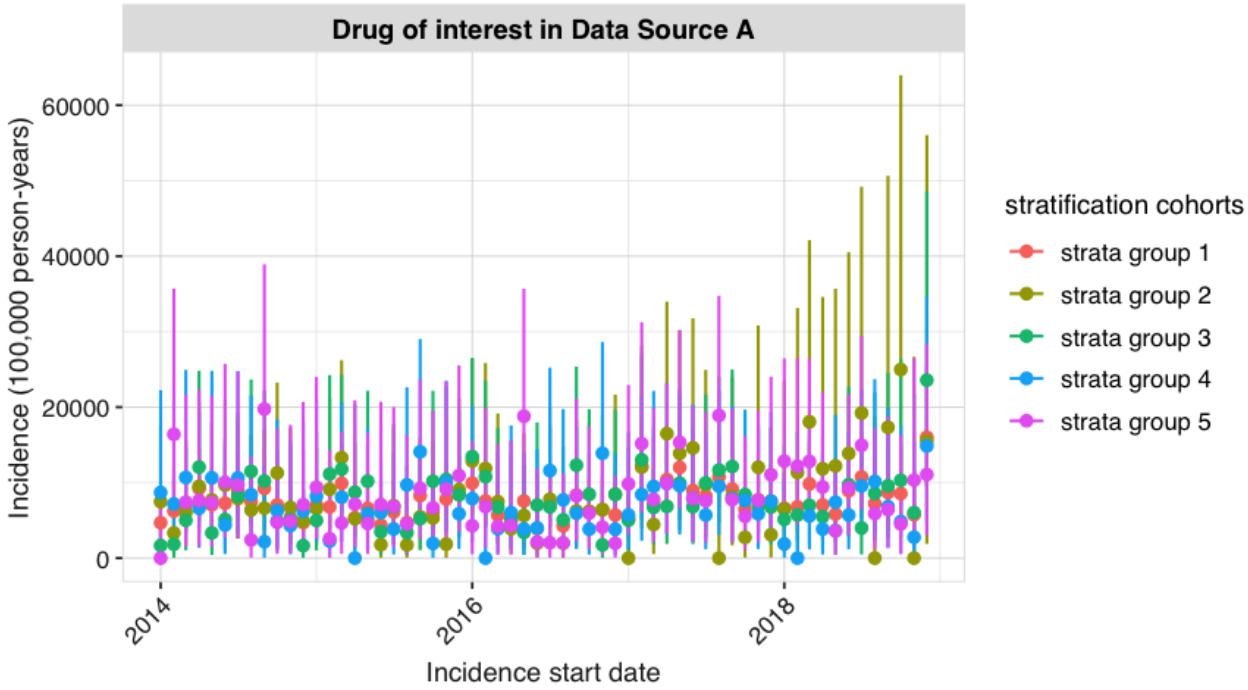


Figure S4. Mock representations of the prevalence plots that will be generated in *Objective 1 – To determine incidence rates estimates*.

Objective 2 – Characterisation of new GLP-1 RA users

2) *To characterise new drug users of GLP-1 RA medicines (overall and stratified by ingredient, by brand, by indication cohorts) by age, sex, initial dose, cumulative dose, and a list of prespecified indications/comorbidities and related co-medications for each data source.*

These analyses will be calculated using incident users (new users of the medicine of interest [e.g., individuals with no prior use of dulaglutide when estimating incidence of dulaglutide use], as shown in **Figure 1B** from **section 9.1.**).

We will generate tables with the characteristics of patients starting a GLP-1 RA medicine for the overall study period and yearly, separately by overall GLP-1 RA use, by ingredient, and by pre-specified brand. Additionally, these will be then stratified by indication as depicted in **Figure S5**.

Results from Objective 2 will provide information of which are the most common characteristics of individuals starting GLP-1 RA, including whether they have an indication for it and potential comorbidities that could be contraindications for the treatment. We will use annual rather than monthly stratification, as monthly estimates are likely to be small and therefore provide limited information. Annual stratification should offer more meaningful insight into changes in the characteristics of new users over time. Changes in the characteristics of new users could indicate shifts in access policy driven by shortages.

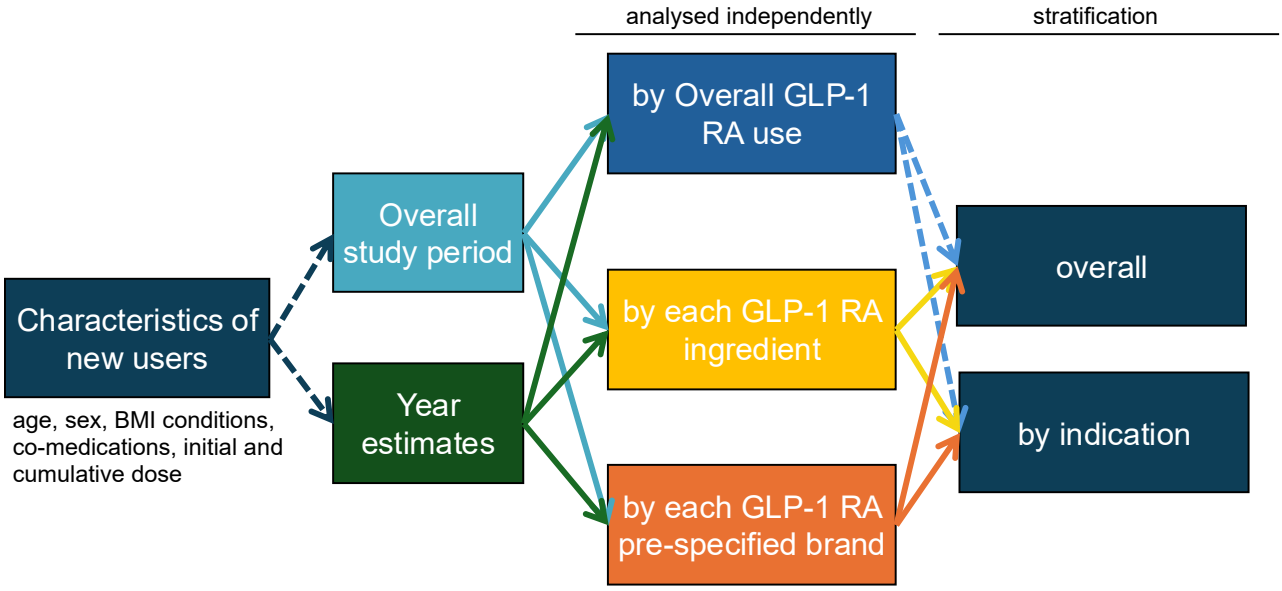


Figure S5. Diagram of the analyses included in *Objective 2 – To characterise new drug users of GLP-1 RA medicines*.

Objective 3 – Switching of new users of any GLP-1 RA users with no prior use of any other GLP-1 RA ingredient in the prior 365 days [overall study period]

3) *To describe GLP-1 RA switching of substances (exenatide, liraglutide, lixisenatide, dulaglutide, semaglutide or tirzepatide cohorts) among new GLP-1 RA users for each data source [overall study period].*

These analyses will be calculated using incident users (new user of any GLP-1 RA medicine with no prior use of any other GLP-1 RA ingredient in the prior 365 days, as shown in **Figure 1B** from **section 9.1.**).

We will generate Sankey diagrams to report the treatments patterns from the first new prescription/dispensation of any GLP-1 RA ingredient 1) to other GLP-1 RA ingredients, 2) to stop GLP-1 RA use, or 3) to one of the pre-defined cohorts (ATC3 and ATC4 level class cohorts described in **Table 4**), during the whole study period, as depicted in **Figure S6**.

Sankey plots will not report time to switch or show differences in treatment patterns over time. To capture time-to-switch, we will report the median time-to-switch and the average time-on-treatment from the first GLP-1 RA using survival analysis in a separate table. To show differences across frequent treatment patterns over time, we might restrict the observation period included in the Sankey diagrams (i.e., stratify by time windows) into 2 time periods (one early, one later). Annual stratifications cannot be included due to excessive number of analysis and combinations that would be computed.

The results from Objective 3 will inform of the most common patterns of treatment at a class level (in other words, it will inform which are the most common switches after starting a GLP-1 RA use with no prior use of any other GLP-1 RA ingredient in the last 365 days). Given the large number of possible combinations that would be required to compute, this study cannot combine the switching of GLP-1 RA ingredients to other GLP-1 RA ingredients and pre-defined cohorts with the strength and brand information. Thus, Objective 4 focuses on switching between GLP-1 RA ingredients with pre-specified strength and brands.

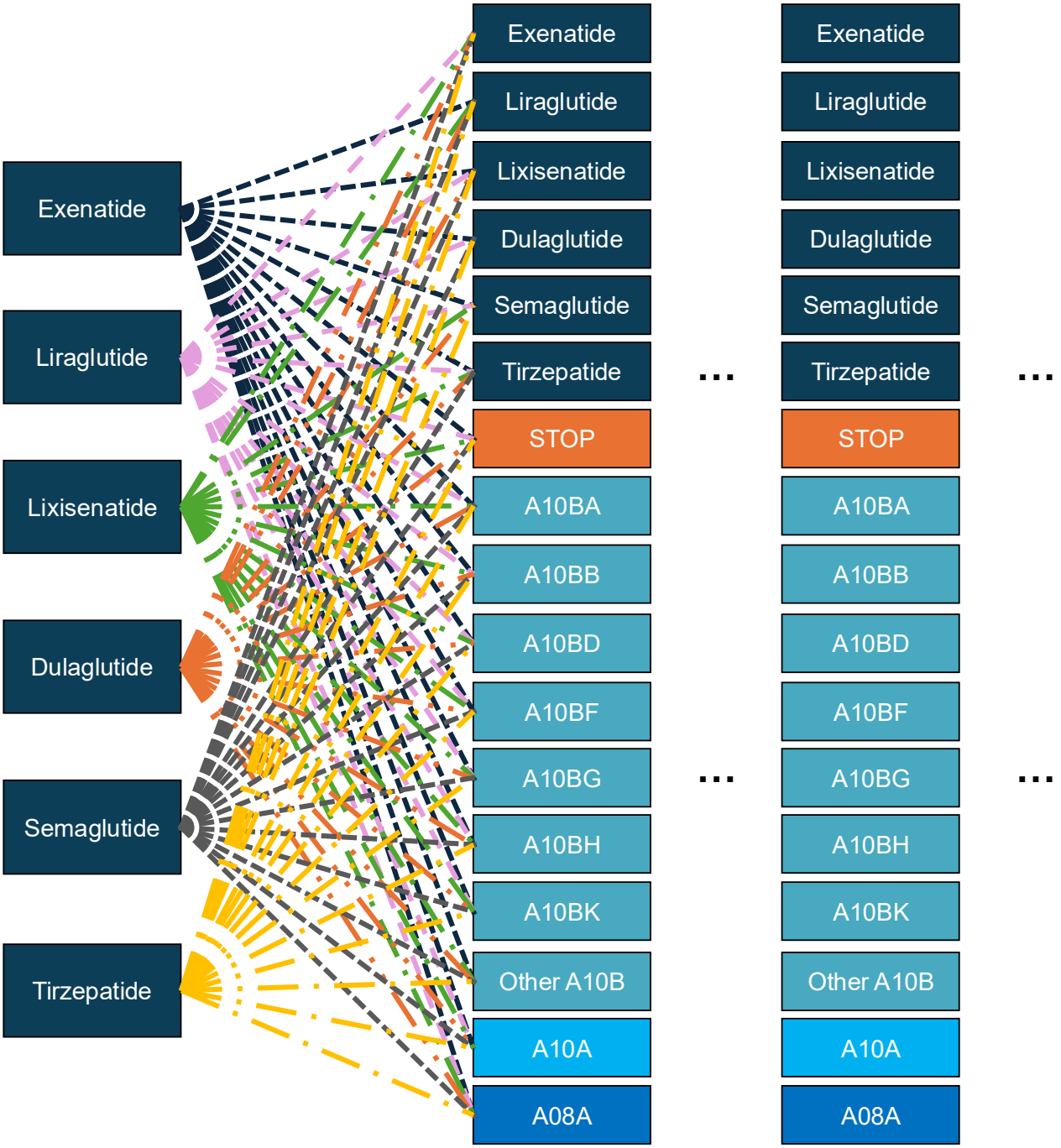


Figure S6. Diagram of the analyses included in Objective 3 – To describe GLP-1 RA switching of substances.

To improve visualisations in the report, we will post-process the results, so the Sankey plots display only the treatment pathways representing more than 5% of individuals in the entire cohort. Full combinations will be displayed in the shiny app. STOP refers to cessation of the last GLP-1 RA treatment received with no switching to any other of the ATC3 and ATC4 level class cohorts.

Abbreviations of ATC3 and ATC4 level class cohorts: A10BA – Biguanides; A10BB – Sulfonylureas; A10BD – Combinations of oral blood glucose lowering drugs; A10BF – Alpha glucosidase inhibitors; A10BG – Thiazolidinediones; A10BH – DPP-4 inhibitors; A10BK – SGLT2 inhibitors; Other A10B – The remaining drugs in A10B class (i.e., excluding GLP-1 RA ingredients and excluding ingredients from the above groups); A10A – Insulin drugs; A08A – Antiobesity preparations, excluding diet products, and excluding A08AA62 (bupropion and naltrexone) and A08AA56 (ephedrine, combinations).

An example of the GLP-1 RA switching plots that will be generated is presented in **Figure S7**.

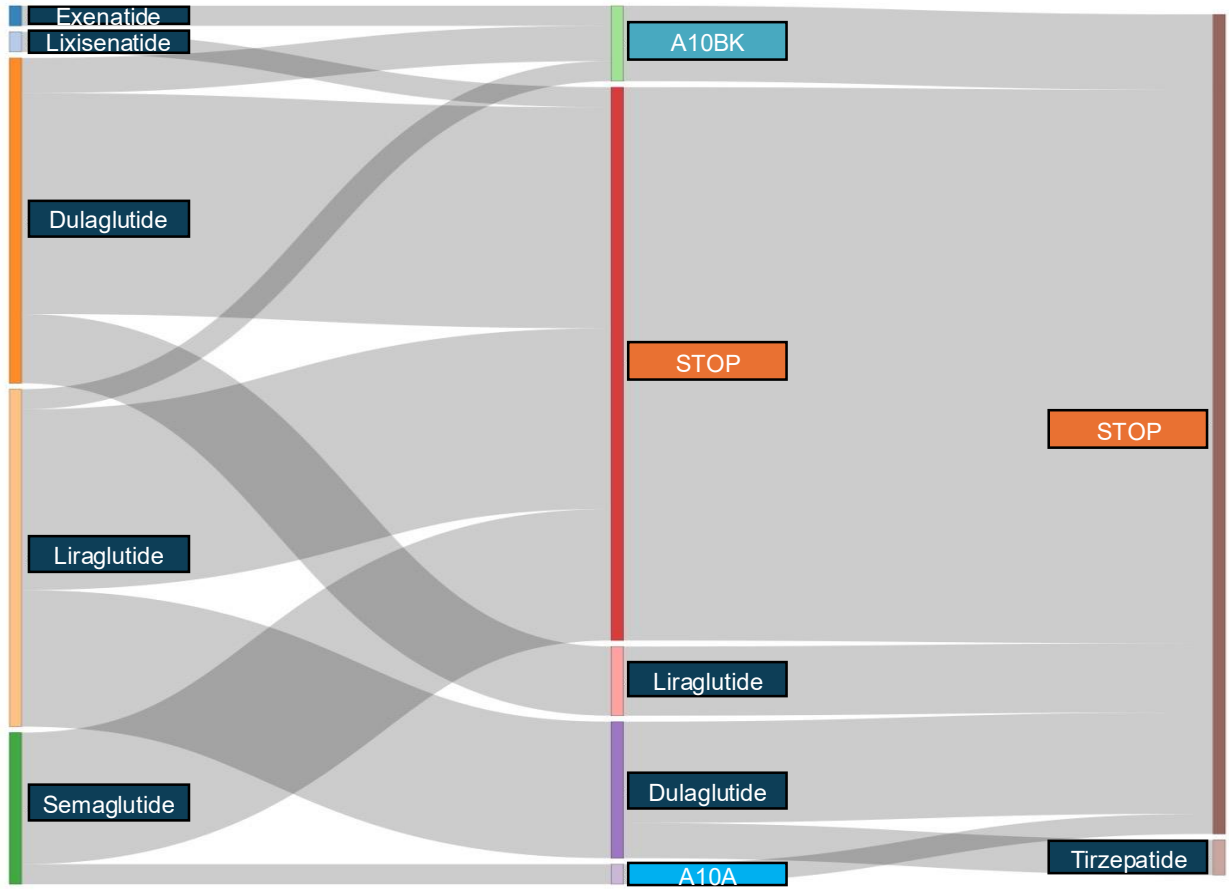


Figure S7. Mock representation of the Sankey plot that will be generated in *Objective 3 – To describe GLP-1 RA switching of substances*.

Abbreviations of ATC3 and ATC4 level class cohorts: A10BK – SGLT2 inhibitors; A10A – Insulin drugs.

Objective 4 – Within-substance switching of new GLP-1 RA users [overall study period]

1) *To describe GLP-1 RA within-substance switching of strength (limited to pre-defined cohorts of strength-levels, with brands when available, for liraglutide, dulaglutide, semaglutide, and tirzepatide) among new users of the respective medicine in each data source [overall study period].*

These analyses will be calculated using incident user (new users of the medicine of interest [e.g., individuals with no use of dulaglutide in the prior 365 days when estimating incidence of dulaglutide use], as shown in **Figure 1B** from **section 9.1.**).

For liraglutide, dulaglutide, semaglutide, and tirzepatide, we will generate separate Sankey diagrams to report the treatment pattern from the first prescription/dispensation of that ingredient 1) to other strengths/brands, or 2) to stop, during the whole study period, as depicted in **Figure S8**.

Sankey plots do not report time-to-switch or show differences in treatment patterns over time. To capture time-to-switch, we will report the median time-to-switch and the average time-on-treatment from the first prescription of the respective substance using survival analysis in a separate table.

The results from Objective 4 will inform of the dose escalation (i.e., switching of strengths) and switching across brands within the same ingredient. Given the large number of possible combinations that would be required to compute, this study cannot combine the switching of strength between different GLP-1 RA ingredients.

A) Pre-defined ingredient cohorts to include in the within-subtance switching

Semaglutide cohorts	
Semaglutide brands	Semaglutide strengths
Wegovy [injection]	0.25mg – Wegovy
	0.5mg – Wegovy
	1mg – Wegovy
	1.7mg – Wegovy
	≥2.4mg – Wegovy
Ozempic [injection]	0.25mg – Ozempic
	0.5mg – Ozempic
	1mg – Ozempic
	≥2mg – Ozempic
Rybelsus [Oral]	1.5mg – Rybelsus
	3mg – Rybelsus
	4mg – Rybelsus
	7mg – Rybelsus
	9mg – Rybelsus
	≥14mg – Rybelsus
No brand recorded for semaglutide	0.25mg – No brand recorded [injection]
	0.5mg – No brand recorded [injection]
	1mg – No brand recorded [injection]
	≥2mg – No brand recorded [injection]
	3mg – No brand recorded [Oral]
	5mg – No brand recorded [Oral]
	7mg – No brand recorded [Oral]
	≥14mg – No brand recorded [Oral]

Liraglutide cohorts	
Liraglutide brands	Liraglutide strengths
Saxenda	0.6mg – Saxenda
	1.2mg – Saxenda
	1.8mg – Saxenda
	2.4mg – Saxenda
	≥3mg – Saxenda
Victoza	0.25mg – Victoza
	0.5mg – Victoza
	≥1.8mg – Victoza
Other brands/no brand recorded for liraglutide	0.6mg – Other brands/no brand recorded
	1.2mg – Other brands/no brand recorded
	1.8mg – Other brands/no brand recorded
	2.4mg – Other brands/no brand recorded
	≥3mg – Other brands/no brand recorded

Dulaglutide cohorts	
Dulaglutide brands	Dulaglutide strengths
Trulicity	0.75mg – Trulicity
	1.5mg – Trulicity
	3mg – Trulicity
	≥4.5mg – Trulicity
No brand recorded for dulaglutide	0.75mg – No brand recorded
	1.5mg – No brand recorded
	3mg – No brand recorded
	≥4.5mg – No brand recorded

Tirzepatide cohorts	
Tirzepatide brands	Tirzepatide strengths
Mounjaro	2.5mg – Mounjaro
	5mg – Mounjaro
	7.5mg – Mounjaro
	10mg – Mounjaro
	12.5mg – Mounjaro
	≥15mg – Mounjaro
No brand recorded for tirzepatide	2.5mg – No brand recorded
	5mg – No brand recorded
	7.5mg – No brand recorded
	10mg – No brand recorded
	12.5mg – No brand recorded
	≥15mg – No brand recorded

B) Example of possible switching within the semaglutide pre-defined cohorts

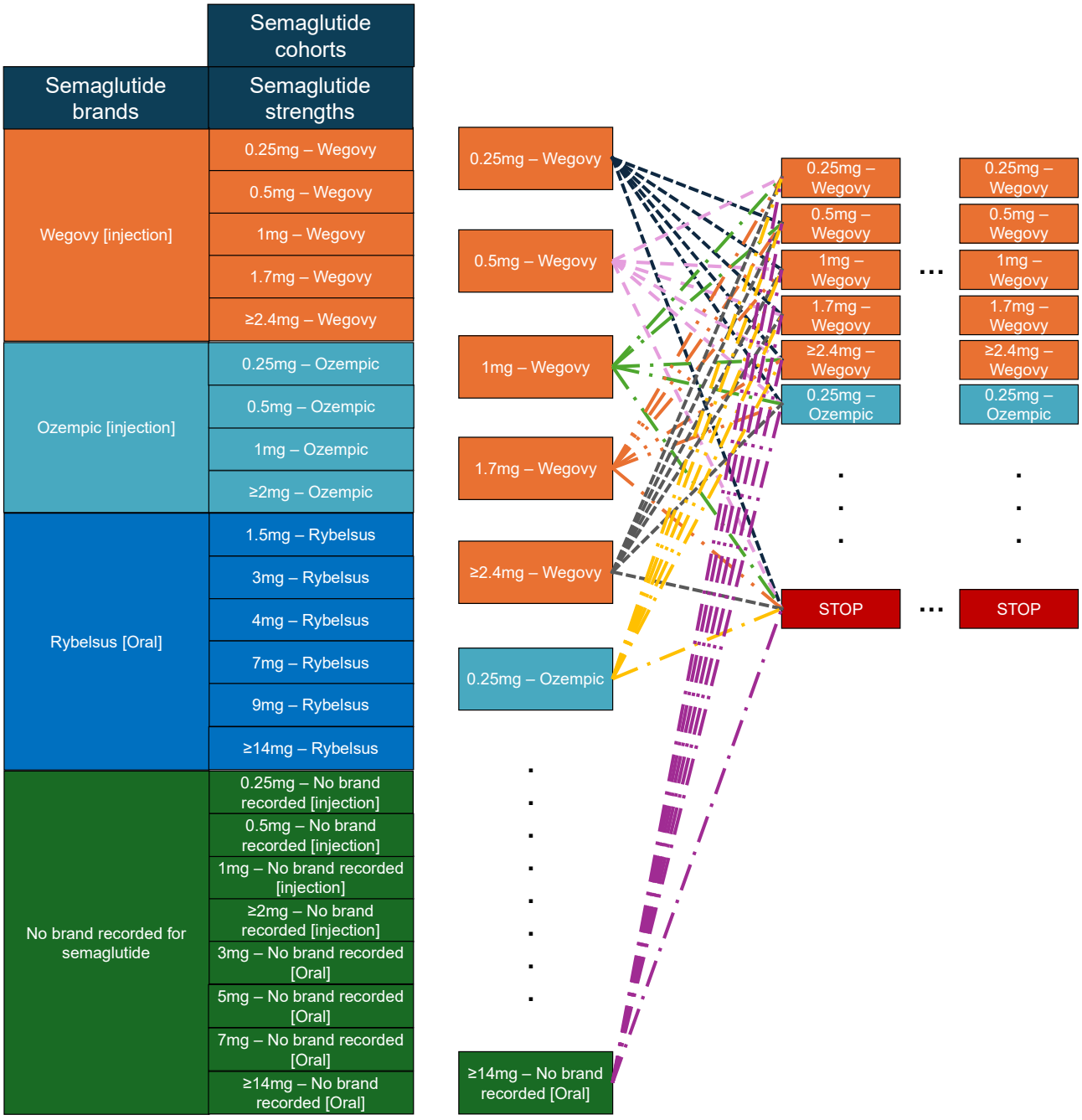


Figure S8. Diagram of the analyses included in *Objective 4 – To describe GLP-1 RA within-substance switching of strength*. A) Pre-defined ingredient cohorts to include in the within-substance switching, and B) example of possible switching within the semaglutide pre-defined cohorts. STOP in includes cessation of the GLP-1 RA treatment (i.e., semaglutide in panel B example) and switch to another GLP-1 RA ingredient.

An example of the GLP-1 RA switching plots that will be generated is presented in **Figure S9**.

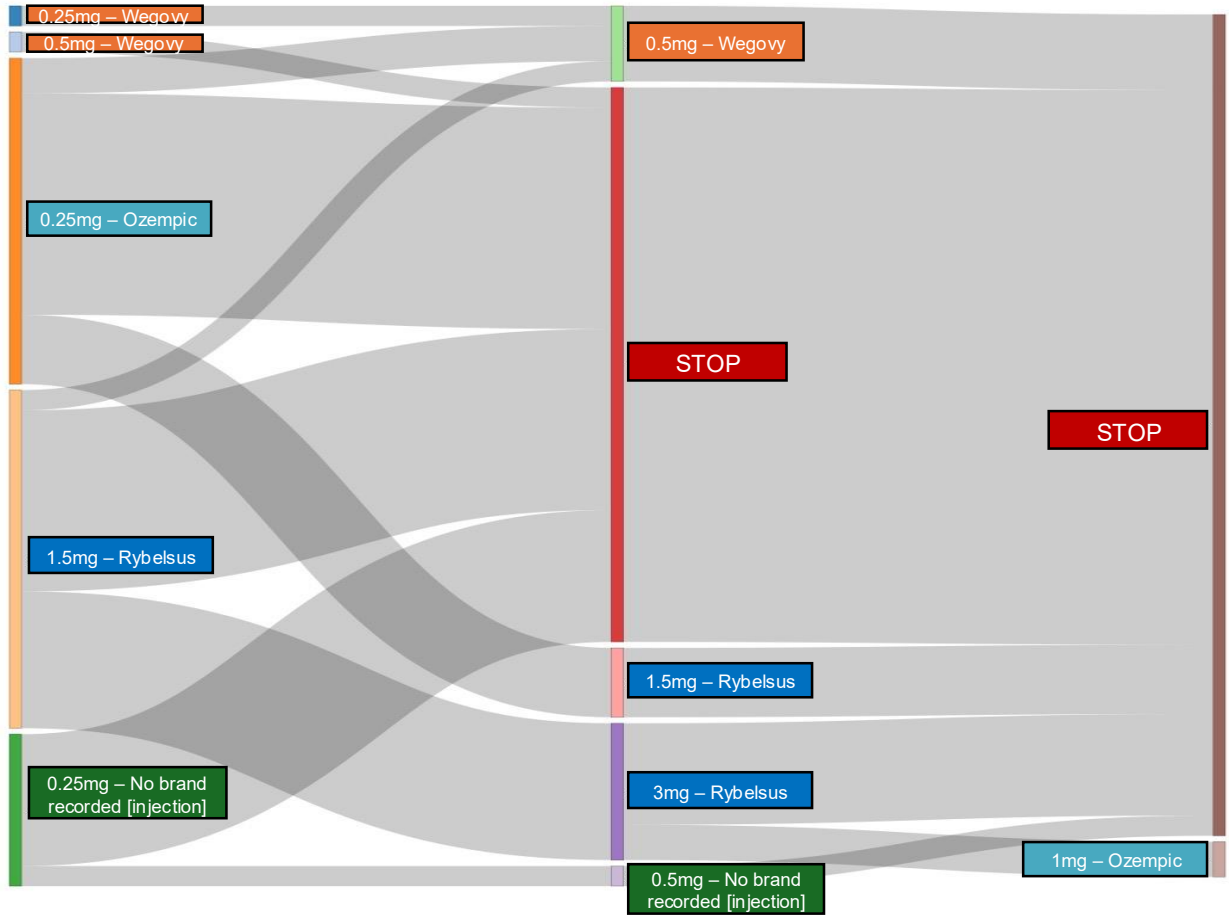


Figure S9. Mock representations of the Sankey plots that will be generated for semaglutide in *Objective 4 – To describe GLP-1 RA within-substance switching of strength*. STOP in includes cessation of the semaglutide treatment and switch to another GLP-1 RA ingredient.

The results from Objectives 3 and 4 can support the design of a complex study focused on switching. By identifying the most frequent patterns, we may be able to exclude ATC3 and ATC4 level class cohorts to reduce the number of potential combinations (i.e., potential patterns that would not need to be computed), which may allow us to merge GLP-1 RA switching of substances and within-substance switching. Additionally, these results will inform about which data sources have sufficient sample size and granularity (i.e., adequate detail on drug strength and brand) to be included in the complex study.

Objective 5 – Switching of new users of any GLP-1 RA users with no prior use of any other GLP-1 RA ingredient in the prior 365 days, and Objective 6 – Within-substance switching of new GLP-1 RA users [between 2015–2020 and annually between 2021 and 2025]

The same analysis described above for Objectives 3 and 4 will be run for different time-period stratifications: between 2015–2020 and annually between 2021 and 2025.

ANNEX II. Description of data sources

DATA SOURCES DESCRIPTION

Belgium: IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)

#	Section	Description
1	Database Identification and country	IQVIA LPD Belgium (IQVIA Longitudinal Patient Database Belgium) Belgium
2	Data partner information section	IQVIA IQVIA Europe
3	Coverage and timespan	Data collection since: 2005 Extent: Nation-wide. Panel of 300 GPs in Belgium. The panel is maintained as a representative sample of the primary care physician population in Belgium, according to three criteria known to influence prescribing: age, sex, and geographical distribution.
4	Healthcare setting / type of data	Primary care – gps. Ambulatory visits, with diagnosis, prescriptions, procedures, and laboratory tests.
5	Data collection process	Outpatient electronic health records. Records are entered by GPs at the healthcare encounter.
6	General representativeness	The panel of contributing physicians (a stable 300 GPs) is maintained as a representative sample of the primary care physician population in Belgium, according to three criteria known to influence prescribing: age, sex, and geographical distribution. The panel consists of a stable 300 GPs that are geographically well spread. The total number of active GPs in Belgium is 15,602. The regional geographical spread of physicians in the LPD data is also representative of the distribution across the country: 57% GPs in the North (compared to 54% nationally), 31% in the South (33% nationally), and 12% in Brussels (13%).The provider of the data has more than 2,250 GPs under contract so in case of a drop out a replacement is easily found.
7	Data content /source coding	No information on source coding.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. The patient ID is per practice. So, a patient can have different IDs in the DB, one per practice. In Belgium, patients are typically registered at only one GP practice, so duplication should be minimal.
9	Quality control (database specific)	No QC. Integrity constraints only.
10	Linkage	No linkage.
11	Vital status	Death information is derived from healthcare events.
12	Limitations	The observation period of the patients in these databases is calculated based on the last visit, observation or interaction of the patient with the health care system. Consequently, healthy and/or non-frequent users of the health care system will not be considered active/included in the

#	Section	Description
		denominator populations during the latest months of available data from the latest data lock. The denominators that will be used to calculate the incident use of drugs in the population may present an artefactual decrease whilst the incident users will remain, incrementing the incidence and prevalence ratios.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111116 Website: https://iqvia.com

Croatia: Croatian National Public Health Information System (NAJS)

#	Section	Description
1	Database Identification and country	NAJS (Croatian National Public Health Information System) Croatia
2	Data partner information section	Croatian Institute of Public Health Department of Data Science and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. Geographic coverage covers whole Croatia, with various levels of resolution for different registries. Current estimates for the population in Croatia will be available at: https://podaci.dzs.hr/hr/podaci/stanovnistvo/procjena-stanovnistva/ for each year.
4	Healthcare setting / type of data	Primary care – gps, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. For both inpatient and outpatient setting diagnoses, medication, procedures, and measurements are captured. Since 2025, also family relationships from the birth registry, histopathological data from the cancer registry, and vital signs from health risk assessment have been added. The year of availability of information depends on the setting: • 2014 for lab tests • 2015 for general practitioners • 2016 for secondary conciliatory care • 2017 for hospital records • 2020 for vaccination records
5	Data collection process	Inpatient hospital billing systems, and Other. Data is entered by clinicians at healthcare contact, then combined by CIPH into the NAJS database.
6	General representativeness	The data is collected from public health records, as the majority of health care in Croatia is public. Personal details are collected to a better extent for insured individuals compared to uninsured patients.
7	Data content /source coding	Medication prescriptions are recorded with ATC codes, and diagnoses with ICD10 codes.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network.

#	Section	Description
		Records from 2017 include insured patients with reliable IDs. Uninsured patients do not have reliable IDs. For example, if a patient changed her status from insured to uninsured, or vice versa, she could be counted several times, as could tracking records from before 2017 and after. By using the unique personal identifier for Croatian citizens, it can be checked and verified.
9	Quality control (database specific)	There is a network of registry personnel (leaders, administrators, coders, sources) working on data coverage and other quality dimensions. An analytical team routinely checks for erroneous entries in hospital records, removing double entries, false dates, and overlapping stays. Entries without enough data or with obviously erroneous dates from primary care analysis are being excluded.
10	Linkage	NAJS is linked to the national death registry, which is updated monthly, and to primary care, which is updated weekly. Additionally, other specific registries are included in NAJS (e.g. diabetes registry), where inclusion criteria vary across these registries.
11	Vital status	NAJS is linked to the national death registry.
12	Limitations	Hospital data is available from 2017 onwards. This is often used as start of data collection, while laboratory and GP data is captured before that (since 2014 and 2015 respectively). The total and active person count in the NAJS data is larger than the current population of Croatia. This explained by a) the person table included deceased and all previously insured people and b) there is no information about insurance ending. It is known that a lot of people migrated (300k-400k) and weren't included in the last population census but still are in the NAJS database. In-hospital administrations are managed via paper drug charts and hospital discharge summaries are currently not captured into NAJS. NAJS do not contain detailed recordings of dose/duration as a substantial proportion of records are mapped at ingredient level, which limits its inclusion in the drug utilisation analyses.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111155 Website: https://www.hzjz.hr/nacionalni-javnozdravstveni-informacijski-sustav-najs/

Denmark: Danish Data Health Registries (DK-DHR)

#	Section	Description
1	Database Identification and country	DK-DHR (Danish Data Health Registries) Denmark
2	Data partner information section	Danish Medicines Agency (DKMA) Data Analytics Centre (DAC)

#	Section	Description
3	Coverage and timespan	Data collection since: 1995 Extent: Nation-wide. The data is representative of the entire Danish population.
4	Healthcare setting / type of data	Community pharmacists, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: diagnosis (including rare diseases and pregnancy data), hospital admissions, discharge and ICU data, Cause of death, Drug prescriptions, dispensing, vaccination and contraception, Procedures, Devices, and Sociodemographic information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. All causes of deaths, all retrieved drug prescriptions, all records of vaccinations, all hospital inpatient and outpatients contacts including disease diagnoses and hospital surgical and non-surgical procedures, cancers, laboratory test results for the entire Danish population from 1/1/1995 onwards.
6	General representativeness	The data is representative of the entire Danish population. Healthcare is free in Denmark, so we do not expect any bias in data collection based on socio-economic status.
7	Data content /source coding	Diagnoses and causes of death are collected using the ICD-10 vocabulary. ATC and RxNorm are used for Drugs. SNOMED codes are used for Procedures.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network.
9	Quality control (database specific)	The data we have received relating to nationwide Danish Health Data registries offer an opportunity for large-scale, population-based studies with several advantages 1) Their large size improves the precision of estimates and enables the study of rare exposures and outcomes with long-term latency, 2) Inclusion of nearly all individuals in the target population ensures that the data reflect routine clinical care and all clinical segments of the source population, 3) Data are collected independently of each research study, thus minimising certain types of bias, e.g., non-response, and the influence from attention to the research question on the diagnostic process. Before the source data is sent to us, the Danish Health Data Authority runs and does comprehensive checks of the registry table data validity of the variables, breaks in data, changes in variable coding, missingness, etc. We perform checks of missingness/completeness in relation to requested variables. In essence, we are receiving a dump of a mirror of the data that is controlled by the SDS. The documentation performed by SDS is available online, in Danish primarily https://www.esundhed.dk/Dokumentation (all variables), but also in English https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/national_health_registers
	Linkage	There is no linkage in this data source.
	Vital status	The Cause of Death registry (DAR) is used, the cause of death is collected using ICD-10 codes.
	Limitations	DK-DHR data source do not have data on specialty, drug brands or strength available. BMI is largely incomplete as weight measurements are only recorded in pregnant women. Record of conditions are limited by individuals who had a hospital contact (i.e., asthma, depression, anxiety and other chronic conditions

#	Section	Description
		that are mainly treated at the GP will likely be captured in individuals who presented severe cases and were hospitalized). However, drug indications are recorded.
13	Main references	Schmidt M, Schmidt SAJ,Adelborg K,Sundbøll J,Laugesen K,Ehrenstein V,Sørensen HT "The Danish health care system and epidemiological research: from health care contacts to database records." Clinical epidemiology (2019): 31372058
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111217 Website: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/healthdatadenmark

Finland: Finnish Care Register for Health Care (FinOMOP - THL)

#	Section	Description
1	Database Identification and country	FinOMOP-THL (Finnish Care Register for Health Care) Finland
2	Data partner information section	Finnish Institute for Health and Welfare (THL) Department of Knowledge Brokers
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. The current CDM population comprises all persons having been alive and residing in Finland since the beginning of 2011.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The THL database covers both public and private, primary, and specialised inpatient and outpatient health care encounters in Finland, starting from 2011. The entire public sector and private inpatient encounters have been included since 2011, while private outpatient encounters, including occupational care, are included since 2020. Since 1998, the register has covered both public outpatient and inpatient specialized care and private inpatient care (TerveysHilmo). Since 2009, the Finnish National Vaccination Register is covered (complete since 2020). The vaccination register covers all vaccinations from the public sector and from a large part of private vaccination providers, with the data coverage from both sections being very good from 2020 onwards. Since 2011, the register has covered public primary care (AvoHilmo). Since 2020, the register has covered private outpatient care and occupational care. In addition, the CDM also contains positive COVID-19 test results from the Finnish National Infectious Diseases Register, which is maintained by THL.
5	Data collection process	Outpatient electronic health records, inpatient hospital electronic health records, and registries. Data is entered by clinicians upon healthcare contact and processed by THL.

#	Section	Description
6	General representativeness	The THL data has national coverage and is therefore representative of the Finnish population. Using the complete population as a basis for the person table also serves to facilitate calculations on a population level, e.g. incidence rates.
7	Data content /source coding	The following coding systems have been OMOP-mapped, typically to a good level of completeness: ICD10fi Finnish Extension, ATC, Toimenpideluokitus (procedure classification adapted from the Nordic Classification of Surgical Procedures (NCSP)), Terveydenhuollon erikoisalat (Hilmo specific provider speciality), Rokotustapa (AR/YDIN National classification for vaccine administration), Tupakointistatus (AR/YDIN National classification for smoking status). Vaccinations are identified on product level based on batch number, trade name, vaccine title, and ATC-code. This is mapped on brand and type in the OMOP CDM.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Each patient in THL has a unique identifier.
9	Quality control (database specific)	The source data collection undergoes a structural and semantic validation before entry into the source database. Additionally, some coded variables undergo quality assessment against the respective code systems post entry into the database. The source registers are also assessed for completeness and coverage, with the aim of improving future collection in the areas where data is lacking.
10	Linkage	THL is already a linkage of multiple Finnish registries (see above).
11	Vital status	The National Population registry data forms the basis for forming the patient population. This ensures an up-to-date location (municipality of residence) of patients, as well as complete death occurrences (although not the cause of death).
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Häkkinen, Pirjo; Mölläri, Kaisa; Saukkonen, Sanna-Mari; Väyrynen, Riikka; Mielikäinen, Lasse; Järvelin, Jutta "Hilmo - Sosiaali- ja terveydenhuollon hoitoilmoitus 2020 : Määrittelyt ja ohjeistus : Voimassa 1.1.2020 alkaen" Terveyden ja hyvinvoinnin laitos (2019):
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111187 Website: https://thl.fi/fi/tilastot-ja-data/ohjeet-tietojen-toimittamiseen/hoitoilmoitusjarjestelma-hilmo

Germany: InGef Research Database (InGef RDB)

#	Section	Description
1	Database Identification and country	InGef RDB (InGef Research Database) Germany
2	Data partner information section	InGef - Institute for Applied Health Research Berlin GmbH

#	Section	Description
3	Coverage and timespan	Data collection since: 2014 Extent: Nation-wide. The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDB currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs. The data in the database depicts all health care use which has been reimbursed by the SHI.
4	Healthcare setting / type of data	Primary care – General Practioner, and community pharmacists, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data. The following data elements are presented in the OMOP CDM: demographic information, diagnoses, procedures, dispensing drugs and advanced therapy medicinal products, vaccinations, pregnancy data (via diagnoses and procedures) and contraception.
5	Data collection process	Insurance/administrative claims. The data in the database depicts all health care use which has been reimbursed by the SHI (statutory health insurances).
6	General representativeness	InGef RDB covers about 11% of the German population and is comparable to the German population in terms of the distribution of age and sex. Most health insurances that contribute to the InGef RDB have nationwide coverage, meaning that the database covers all regions of Germany. Since almost all services covered by statutory health insurances are specified in national legislation, healthcare provision all over Germany is well represented in the InGef RDB. Additionally, in Germany it is very common to stay with the same health insurance throughout life, which results in a good longitudinal coverage over the entire period of 10 years.
7	Data content /source coding	The coding in the research database complies with national classification and coding rules in Germany. Diagnoses are coded according to ICD-10-GM. Inpatient and outpatient surgeries or procedures are recorded as OPS codes (German classification of Operations and Procedures). The dispensing of drugs in pharmacies is recorded using the PZN (pharmaceutical registration number). For drugs that miss a PZN-to-RxNorm mapping, the ATC code is used instead. In some cases, dispensed drugs can be coded using OPS codes (e.g. in hospitals) or EBM codes (fee schedule for outpatient treatments).
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In the German statutory health system, a person can only be enrolled in one health insurance at a time. However, if a person changes from one contributing insurer to another, a new ID number will be generated.
9	Quality control (database specific)	The data transmitted by healthcare providers complies with the standardized requirements and formats of the Association of Statutory Health Insurances (GKV-SV). Before being imported into the research database, the data elements are checked for data format, completeness and plausibility. After each update of the research database, various counts are compared with the previous update to verify completeness. Due to the

#	Section	Description
		anonymity of the database, direct validation of the data (e.g., using medical records as the gold standard) is not possible.
10	Linkage	Due to the anonymization of the source data, linkage is not possible.
11	Vital status	The date of death is recorded as the last day of the quarter in which the death occurred (i.e. 30/31st of Mar/Jun/Sept/Dec) as reported to the health insurance (no linkage to death registry). The cause of death is not available.
12	Limitations	Ambulatory diagnoses and procedures are summarised in the source on a quarterly basis. Both are mapped to the observation table with the date set to the last day of the respective quarter (i.e. 30/31st of Mar/Jun/Sept/Dec) and the concept "History of event within 3 months" (observation_concept_id 1340222), with the actual diagnosis or procedure concept_id recorded in the field "value_as_concept_id". There is no vocabulary for the German pharmaceutical product codes (PZN). A direct source-to-standard-mapping has been done manually by InGef RDB but is incomplete. The drug exposure duration is unknown. Following OMOP conventions, the end date is always set to dispensing date + 29. Outpatient and inpatient procedures are recorded as OPS codes (German Procedure Classification), for which the vocabulary is incomplete. Approx. 10.5 Million insurees are included in the database, 7.8 Million of these actively insured in 2024. This corresponds to 7% of the total German population. Data are longitudinally linked over a period of currently ten years.
13	Main references	Andersohn F, Walker J "Characteristics and external validity of the German Health Risk Institute (HRI) Database." Pharmacoepidemiology and drug safety (2016): 26530279
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111207 Website: https://www.ingef.de/en/

Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)

#	Section	Description
1	Database Identification and country	IQVIA DA Germany (IQVIA Disease Analyzer Germany) Germany
2	Data partner information section	IQVIA
3	Coverage and timespan	Data collection since: 1989 Extent: Nation-wide. GP and specialists in Germany using specific patient management software.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians). Diagnoses, medication, and procedures from an ambulatory setting. Medications are recorded as prescriptions of marketed products.
5	Data collection process	Outpatient electronic health records. By clinicians at healthcare contact.
6	General representativeness	No specific details on general representativeness given.

#	Section	Description
7	Data content /source coding	Prescription is on product code level (German PZN), ICD10, NFC, Local lab coding.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. There can be patients registered under different ID numbers, because there is no linkage between different GPs.
9	Quality control (database specific)	Data is quality checked on plausibility.
10	Linkage	No.
11	Vital status	Death information is derived from medical events.
12	Limitations	The observation period of the patients in these data sources is calculated based on the last visit, observation or interaction of the patient with the health care system. Consequently, healthy and/or non-frequent users of the health care system will not be considered active/included in the denominator populations during the latest months of available data from the latest data lock. The denominators that will be used to calculate the incident use of drugs in the population may present an artefactual decrease whilst the incident users will remain, incrementing the incidence and prevalence ratios.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104282 Website: https://www.iqvia.com/

The Netherlands: Integrated Primary Care Information (IPCI)

#	Section	Description
1	Database Identification and country	IPCI (Integrated Primary Care Information) The Netherlands
2	Data partner information section	Erasmus University Medical Center Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards. However, it mainly covers the central part of the country, including the most densely populated area (the 'Randstad') and non-urban areas. IPCI contains information on all patients registered with GPs responsible for non-emergency care and referrals. A patient is registered at birth or at first encounter with the GP.
4	Healthcare setting / type of data	Primary care – gps. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.

#	Section	Description
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. Data is aggregated by Erasmus MC data managers and combined in one harmonized database. Several checks are done on this database to ensure correct data processing. Persons are mostly uniquely identified, with the exception of when persons change GP practice (when the same individual can receive several different identifiers).
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content /source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients can be registered under different IDs. However, in the Netherlands, patients typically have one GP and changing practice is uncommon.
9	Quality control (database specific)	Prior to each data release, extensive quality control steps are performed, e.g., comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g. number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical data, e.g. distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.

#	Section	Description
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.
13	Main references	de Ridder MAJ, de Wilde M,de Ben C,Leyba AR,Mosseveld BMT,Verhamme KMC,van der Lei J,Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182143
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

Norway: Norwegian Linked Health Registry data (NLHR)

#	Section	Description
1	Database Identification and country	NLHR (Norwegian Linked Health Registry data) Norway
2	Data partner information section	University of Oslo Faculty of Mathematics and Natural Science – Department of Pharmacy
3	Coverage and timespan	Different registries have different timespans: National registry to identify the population: 2018 – 2023 Diagnosis from secondary care (incl. history of disease): since 2008 Extent: Nation-wide. Norway has a universal public health care system, consisting of primary and specialist health care services covering a population of approximately 5.4 million inhabitants.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following registries (relevant in this study) are included: the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Surveillance System for Communicable Diseases (MSIS), and the National Registry (NR).
5	Data collection process	Registries. Many population-based health registries were established in the 1960s, with use of unique personal identifiers facilitating linkage between registries. Data in these health registries are used for health analysis, health statistics, improving the quality of healthcare, research, administration, and emergency preparedness.
6	General representativeness	The NLHR data covers the full Norwegian population.
7	Data content /source coding	NPR: ICD-10 for diagnosis, ATC and some special codes for drug use, Norwegian codes for clinical procedures (surgery (NCSP), medicine (NCMP) and diagnostic imaging, image-guided intervention, and nuclear medicine

#	Section	Description
		(NCRP)). KUHR: ICD-10 and ICPC-2 and ICPC-2B for diagnosis/procedure. NorPD: ATC.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Linkage between the registries was facilitated using project-specific person IDs generated from unique personal identification assigned at birth or immigration for all legal residents in Norway.
9	Quality control (database specific)	In-house data quality checks of rates of common conditions, drug exposures, and outcomes. We compare obtained rates with official national statistics (e.g., birth statistics, yearly rates of drug dispensing, and diagnosis by age and gender). We also review missing data and outliers and inform registry holders of any unusual patterns.
10	Linkage	The NLHR is, by definition, a linkage of datasets. Helsedata.no is one central portal to apply for 11 national health registries, including all the registries that have been mapped to the OMOP CDM.
11	Vital status	The national death registry is linked.
12	Limitations	NLHR data sources do not have data on drug brands or strength available. BMI is largely incomplete as weight measurements are only recorded in pregnant women at the start of pregnancy. Primary care data in NLHR starts from 2008, while drug data and secondary care data start from 2018, and in all of datasets are available to the end of 2023.
13	Main references	Trinh NT, et al. Harmonizing Norwegian registries onto OMOP common data model: Mapping challenges and opportunities for pregnancy and COVID-19 research. *Int J Med Inform*. 2024;191:105602. doi:10.1016/j.ijmedinf.2024.105602
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000409 Website: https://www.mn.uio.no/farmasi/english/research/groups/pharma-safe/

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

#	Section	Description
1	Database Identification and country	SIDIAP (The Information System for the Development of Research in Primary Care) Catalunya, Spain
2	Data partner information section	IDIAPJGol
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. The SIDIAP database contains records of around 6 million people residing in Catalonia, estimated to be representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – gps, and hospital inpatient care. SIDIAP captured data includes routine visits, demographics, diagnoses,

#	Section	Description
		laboratory tests, drugs (prescribed and dispensed), referrals, and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, among other. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Catala de la Salut is the owner of the data and acts as the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content /source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network.
9	Quality control (database specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenised (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns of a certain variable in the registry. With this information, the SIDIAP team can issue recommendations to researchers about the which variables are likely to record certain information (e.g., there are several variables with information concerning the women's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinally and reliability), plausibility (range checks and unusual values), and consistency are inspected through visualisation tools. In addition, before accessing the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences, and prevalence of the data requested for the project, allowing detection of inconsistencies in the data extraction prior to data delivery. External validation processes of the SIDIAP database mainly include assessing the data recorded in SIDIAP through linkage to external gold

#	Section	Description
		standard data sources, by analysing free text, or by sending questionnaires to health professionals.
10	Linkage	SIDIAP is linked to a hospital discharge database, pharmacy dispensation, and primary care laboratories. It can also be linked to other registries in Catalonia on a project-by-project basis.
11	Vital status	Mortality is fully captured in SIDIAP. The cause of death is not available but can be linked to the Spanish death registry on a project-by-project basis.
12	Limitations	The SIDIAP data is not representative of individuals not using public primary care, and conditions that are usually followed by specialist care might not be properly captured. In addition, there is limited information on lifestyle variables. Patients are followed until Death or when transferring to another primary health care centre that does not contribute to SIDIAP.
13	Main references	Recalde M, Rodríguez C,Burn E,Far M,García D,Carrere-Molina J,Benítez M,Moleras A,Pistillo A,Bolíbar B,Aragón M,Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

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

#	Section	Description
1	Database Identification and country	HI-SPEED (Health Impact - Swedish Population Evidence Enabling Data-linkage) Sweden
2	Data partner information section	SMPA-GU, Läkemedelsverket, Box 26 Pharmacoepidemiology and Analysis Department (FeA)
3	Coverage and timespan	Data collection since: 2020 Extent: Nation-wide. The catchment area includes the whole of Sweden, covering the full population of approximately 10 million.
4	Healthcare setting / type of data	Primary care – gps, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: Socio-demographics, drug use and prescriptions, diagnoses, cause of death, primary care procedures and visits, as well as secondary care and inpatient visits or clinical events.
5	Data collection process	Registries. The data is acquired from the Swedish nation and regional registries, only once all legislative, GDPR and ethical approvals have been granted. Therefore, only relevant data is passed on, which will then be entered and processed by the study team. The data are updated several times annually.
6	General representativeness	The coverage includes all patients of all sociodemographic characteristics. Therefore, it should mirror the source population to a very good extent.
7	Data content /source coding	Medicines are coded with ATC, ICD10 is used for diagnoses, and the Swedish procedure coding system (KVA) is used for clinical procedures.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and

#	Section	Description
		semantic conformance has been verified upon onboarding into the DARWIN EU® data network.
9	Quality control (database specific)	The source data is obtained from the Swedish National and Regional Registers. The registers perform some regular quality controls on their data. After receiving the data, we perform additional checks and cleaning. We also run regular quality checks on the data we manage.
10	Linkage	Data on specialist care is acquired from the National Patient Register; mortality information is provided by the Cause-Of-Death Registry; drugdata is provided by the Patient Drug Register.
11	Vital status	Data on death and cause-of-death are extracted from the Cause-of-Death registry.
12	Limitations	Study period in HI-SPEED starts in 2018. Due to this data availability, incidence rates will be restricted to 2019–onwards.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/node/4463/ Website: https://www.gu.se/en/research/scifi-pearl

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

#	Section	Description
1	Database Identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) The United Kingdom
2	Data partner information section	University of Oxford NDORMS
3	Coverage and timespan	Data collection since: 1987 Extent: Nation-wide. CPRD GOLD consists of patients in contributing practices using Vision software. Historically this covered the whole of the UK, but the number of contributing practices in the England is dropping. In January 2025 only 3 practices from England were a part of CPRD GOLD, while historical patient data were from the whole of the UK, and will continue to be so. In the future, no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records. Data is entered by clinicians into the EHR. Data is processed by CPRD and provides data releases for research.
6	General representativeness	CPRD GOLD has been assessed and found to be broadly representative of the UK general population in terms of age, gender, and ethnicity. In CPRD GOLD in January 2025 there were 2,730,707 current acceptable patients (i.e. registered at currently contributing practices that use Vision software,

#	Section	Description
		excluding transferred out, deceased patients, and those flagged by CPRD as not acceptable for clinical research for data quality issues). This equals to 4.07%, based on the UK population estimates of 67,026,300 from the Office of National Statistics (mid-2023). Current patients are only from Scotland, Wales, and Northern Ireland. Historically, GOLD does contain data from England as well.
7	Data content /source coding	Gemscript, Read, dm+d
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In GOLD, a patient can be registered under different ID numbers upon changing practice or re-registration. Researchers are not able to identify these patients, as the data are anonymised. However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (database specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (uts). Each practice is associated to an uts date set when the data quality standards become satisfactory, and CPRD recommend using only longitudinal data starting from this uts date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice uts date. CPRD also check data quality standards at the patient level and associate each patient to a flag, reporting if its data is acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at Practice Level
11	Vital status	Vital status is retrieved from the GP records. Population registry (ONS) data can be requested on a study-by-study basis and linked. This data only covers England and is planned to be mapped to OMOP in the future. The cause of death is not captured.
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL,Dedman D,Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://cprd.com

ANNEX III. Additional information

DATA MANAGEMENT

Data management

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

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

Data storage and protection

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

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

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

QUALITY CONTROL

General data source quality control

A number of open-source quality control mechanisms for the OMOP CDM have been developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). In particular, it is expected that data partners will have run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides numerous checks relating to the conformance, completeness, and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions; completeness in the sense of data quality is solely focused on quantifying missingness, or the absence of data; finally, 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 drug cohorts, non-systemic products will be excluded from the list of included codes summarised on the ingredient level. A pharmacist will review the codes of the drugs of interest.

When defining cohorts for indications, a systematic search of possible codes for inclusion will be identified using *CodelistGenerator* R package (https://github.com/darwin-eu/CodelistGenerator). This software allows the user to define a search strategy and using this will then query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, *PhenotypeR* (https://github.com/OHDSI/PhenotypeR) and *DrugExposureDiagnostics* R package (https://github.com/darwin-eu/DrugExposureDiagnostics) will be run as diagnostics 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 two R packages currently being developed to (1) estimate Incidence and Prevalence and (2) characterise drug utilisation using the OMOP common data model. These packages will include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub.

PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

A PDF report including an executive summary, and the specified tables and/or figures will be submitted to EMA by the DARWIN EU® CC upon completion of the study.

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

ANNEX IV: List of preliminary concept IDs

Table S1. Preliminary list of medicines definitions.

Substance/ATC group Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
Exenatide	Exenatide	Ingredient	A10BJ01	1583722	Yes
Liraglutide	Liraglutide	Ingredient	A10BJ02	40170911	Yes
Lixisenatide	Lixisenatide	Ingredient	A10BJ03	44506754	Yes
Dulaglutide	Dulaglutide	Ingredient	A10BJ05	45774435	Yes
Semaglutide	Semaglutide	Ingredient	A10BJ06	793143	Yes
Tirzepatide	Tirzepatide	Ingredient	A10BX16	779705	Yes
Semaglutide	Wegovy	Brand Name	A10BJ06	TBD	TBD
Semaglutide	Ozempic	Brand Name	A10BJ06	TBD	TBD
Semaglutide	Rybelsus	Brand Name	A10BJ06	TBD	TBD
Liraglutide	Victoza	Brand Name	A10BJ02	TBD	TBD
Liraglutide	Saxenda	Brand Name	A10BJ02	TBD	TBD
Tirzepatide	Mounjaro	Brand Name	A10BX16	TBD	TBD
Biguanides	Biguanides	ATC 4th	A10BA	21600745	YES
Sulfonylureas	Sulfonylureas	ATC 4th	A10BB	21600749	YES
Combinations of oral blood glucose lowering drugs	Combinations of oral blood glucose lowering drugs	ATC 4th	A10BD	21600765	YES
Alpha glucosidase inhibitors	Alpha glucosidase inhibitors	ATC 4th	A10BF	21600775	YES
Thiazolidinediones	Thiazolidinediones	ATC 4th	A10BG	21600779	YES
DPP-4 inhibitors	Dipeptidyl peptidase 4 (DPP-4) inhibitors	ATC 4th	A10BH	21600783	YES
SGLT2 inhibitors	Sodium-glucose co-transporter 2 (SGLT2) inhibitors	ATC 4th	A10BK	1123627	YES
Other A10B class (excluding above substances)	Other blood glucose lowering drugs, excl. insulins; Sulfonamides (heterocyclic)	ATC 4th	A10BX; A10BC	21600788; 21600763	YES
Insulin drugs	INSULINS AND ANALOGUES	ATC 3rd	A10A	21600713	YES
Antiobesity preparations, excluding diet products	ANTIOBESITY PREPARATIONS, EXCL. DIET PRODUCTS	ATC 3rd	A08A	21600680	YES

TBD = to be determined.

Table S2. Preliminary list of conditions from frailty index.

Condition	Concept ID	Concept name
Mobility and transfer problems	4053076	Mobility poor
	4306934	Impaired mobility
	4052049	Mobility fair
	4310235	Reduced mobility
	4031883	Impaired bed mobility
	4032531	Impaired wheelchair mobility
	4052047	Mobility very poor
	4119464	Does not transfer between wheelchair and toilet
	4118805	Unable to transfer between wheelchair and toilet
	4199114	Difficulty mobilizing using mobility aids
	4199113	Does not mobilize using mobility aids
	4199111	Unable to mobilize using mobility aids
	4023190	Wheelchair bound
	4136754	Dependent on helper pushing wheelchair
	4200353	Able to mobilize using mobility aids
	4199115	Able to mobilize using wheelchair
	4199721	Able to move around supporting self on furniture
	44790310	Able to walk short distances
	45878557	Completely immobile
	36716239	Dependent for sitting
	36716240	Dependent for standing
	4146424	Dependent for walking
	46272933	Deterioration in ability to walk
	4200194	Difficulty mobilizing
	4199094	Difficulty mobilizing indoors
	4199114	Difficulty mobilizing using mobility aids
	4199116	Difficulty mobilizing using wheelchair
	4199431	Difficulty moving
	4200817	Difficulty moving around supporting self on furniture
	4107789	Difficulty shuffling
	4107851	Difficulty sitting
	4199552	Difficulty sitting unsupported
	4093668	Difficulty standing
	4154006	Difficulty transferring weight

Condition	Concept ID	Concept name
	36714126	Difficulty walking
	4086550	Difficulty walking up a slope
	4112788	Difficulty weight-bearing
	4199112	Does mobilize using aids
	4199725	Does mobilize using wheelchair
	4200815	Does move around supporting self on furniture
	4200193	Does not mobilize
	4200798	Does not mobilize indoors
	4200183	Does not move
	4084746	Does not shuffle
	4106333	Does not sit
	4199551	Does not sit unsupported
	4106335	Does not stand
	4154005	Does not transfer weight
	4086871	Does not walk
	4086549	Does not walk up a slope
	4112787	Does not weight-bear
	4295037	Get up and go test - abnormal
	4009877	Immobile
	1621081	Immobile or
	4031883	Impaired bed mobility
	4306934	Impaired mobility
	3198828	Increased weakness when ambulating
	4010359	Loss of control of walking
	1314392	Patient is not ambulatory, bed ridden, immobile, confined to chair, wheelchair bound, dependent on helper pushing wheelchair, independent in wheelchair or minimal help in wheelchair
	1314394	Patient not ambulatory, bed ridden, immobile, confined to chair, wheelchair bound, dependent on helper pushing wheelchair, independent in wheelchair or minimal help in wheelchair
	44790681	Patient unable to get up unaided
	4012646	Stick only for walking
	4012944	Tripod/quadrupod: walking
	4199550	Unable to mobilize
	4199093	Unable to mobilize indoors

Condition	Concept ID	Concept name
	4199111	Unable to mobilize using mobility aids
	4200355	Unable to mobilize using wheelchair
	4200350	Unable to move around supporting self on furniture
	4105451	Unable to shuffle
	4106332	Unable to sit
	4023187	Unable to sit unsupported
	4060223	Unable to stand
	4151066	Unable to transfer weight
	4086548	Unable to walk
	44792042	Unable to walk long distances
	4086874	Unable to walk up a slope
	4116707	Unable to weight-bear
	44789400	Uses wheelchair outdoors
	4012945	Uses zimmer frame
	45878235	Very limited, Immobile
	4266144	Walking aid use - finding
	439405	Walking disability
	4240470	wheelchair
	4086557	Difficulty walking up stairs
	4200822	Difficulty managing stairs
	4022073	Dependence on wheelchair
Housebound	40299189	Housebound
	4052962	Housebound
	45877743	Bedridden
	4022076	Patient dependence on care provider
	4019828	Personal care assistance at home (procedure)
	4022523	General home assistance of patient
Activity limitation	44811145	Unfit for activity
	5767124	Difficulty performing personal grooming activity
	36716238	Dependent for personal hygiene activity
	4110470	Difficulty performing personal hygiene activity
	4109859	Unable to perform bathing activity
	4032520	Activity of daily living (ADL) alteration
	4031882	Activity alteration
	4137049	Disability affecting daily living

Condition	Concept ID	Concept name
	4030753	Physical functional dependency (finding)
	36713755	Functionally dependent
Visual impairment	4265433	Visual impairment
	4023310	Blindness AND/OR vision impairment level (disorder)
	40305578	Blindness
	44797518	Visual disturbances and blindness
	375545	cataract
	37541	Glaucoma
	374034	Visual disturbance
	42872584	Registration of visual impairment
Hearing impairment	36715579	Acquired hearing loss
	377889	Hearing loss
	439378	Ear anomalies with hearing impairment
	444291	Sensory hearing loss
	44805060	Wears bone anchored hearing aid
	379832	Mixed conductive AND sensorineural hearing loss
	378444	Hearing disroder
	42539697	External hearing aid in situ
	4246497	Hearing aid
Requirement for care	4192880	Lives in a residential home
	35609081	Home visit requested by care home staff
	4052486	Lives in a nursing home
	4192880	Lives in a residential home
	44791364	Lives in care home
	4074789	Lives in supported home
	4022081	Living in residential institution
	3661927	Living temporarily in care home
	44790305	Local authority care home
	40486978	Nursing home acquired pressure ulcer
	44790706	Pain commenced - residential home
	44791204	Place of occurrence of injury: residential home environment
	44802299	Previously lived in care home
	44788859	Private or voluntary care home
	4147552	Private residential home
	765265	Problem related to living in residential institution

Condition	Concept ID	Concept name
	36713971	Referred by care home
	44814152	Referred by nursing home
	44814153	Referred by residential home
	4119866	Residential home
	44804659	Residential home acquired pressure ulcer
	4305680	Residential institution
	37310422	Seen by clinical pharmacist in care home
	4088536	Seen in nursing home
	44788849	Provision of residential care (regime/therapy)
	37108723	Assisted living facility patient
	44807727	Has a paid carer
	42535090	Need for personal care assistance
	4081589	Has a caregiver
Social vulnerability	4019836	Social exclusion
	4228687	Impaired social interaction
	4309238	Social isolation
	4019835	Social withdrawal
	4297462	Social isolation (rejection)
	4172829	Limited social contact
	4317527	Family-related social factor
	42690410	Carer behavior is cause for safeguarding concern
	44792191	Extensive support provided by carer
	4147192	Feeling lonely
	44805255	Has an informal carer
	44805672	Has an older carer
	44807727	Has a paid carer
	44805674	Has a parent carer
	37394063	Has kinship carer
	44813864	Has socially isolated carer
	44806914	Has voluntary carer
	44792192	Inadequate support provided by carer
	4023168	Lives alone
	4052158	Lives alone needs housekeeper
	4053087	Lives alone no help available
	45879223	Lonely

Condition	Concept ID	Concept name
	44789099	Parent is informal carer
	44789487	Partner is informal carer
	44810043	Referred by Social Services
	44790469	Relative is informal carer
	44789986	Report received from social services
	37208707	Requires carer to be present at encounters
	4052789	Social problem
	4209159	Social problem not due to a mental disorder
	44788883	Under care of social services
	4221049	Vulnerable adult
	44791055	Vulnerable elderly person
	44791931	Vulnerable family
	4151777	Vulnerable family support
	44803964	Vulnerable group
	4116985	Vulnerable personality
Falls	4087528	Recurrent falls
	4256754	Falls caused by medication
	4224116	Unexplained recurrent falls
	4329906	At risk for injury due to fall
	436583	Fall (observation)
	435991	accidental fall
	4184243	Elderly fall
	4323345	History of fall
Urinary incontinence	193326	Urge incontinence of urine
	193598	Extravasation of urine
	193874	Nocturnal enuresis
	195007	Female stress incontinence
	195079	Functional urinary incontinence
	197102	Unaware of passing urine
	197378	Overflow incontinence of urine
	197672	Urinary incontinence
	443524	Mixed urinary incontinence
	444035	Incontinence
	606405	Extra urethral urinary incontinence
	606955	Stress incontinence following surgical procedure

Condition	Concept ID	Concept name
	4012368	Increased frequency of urination
	4030763	At risk for urge incontinence
	4032498	Abnormal bladder continence
	4032530	Total urinary incontinence
	4092642	Urinary loss
	4096552	Unaware of need to urinate
	4126278	Postural urinary incontinence
	4153667	Urinary incontinence due to urethral sphincter incompetence
	4172646	Urinary incontinence of non-organic origin
	4302457	Double incontinence
	4314023	Incontinence due to detrusor instability
	37119132	Urinary incontinence co-occurrent and due to prolapse of female genital organ
	37208161	Daily urinary incontinence
	40480232	Male urinary stress incontinence
	40481801	Stress incontinence after prostatectomy
	40490423	Incontinence without sensory awareness
	42536555	Stress incontinence co-occurrent and due to pelvic organ prolapse
	42538537	Overflow incontinence of urine due to prolapse of female genital organ
	42538538	Urge incontinence due to prolapse of female genital organ
	42538539	Mixed incontinence due to prolapse of female genital organ
	42872846	Intermittent urinary incontinence
	44808460	Sneezing incontinence of urine
	45757352	Urinary incontinence due to benign prostatic hypertrophy
	45770268	Functional urinary and faecal incontinence
Memory and cognitive problems	439795	Minimal cognitive impairment
	443432	Impaired cognition
	761978	Cognitive impairment due to multiple sclerosis
	3654469	Amnestic mild cognitive disorder
	3654907	Cognitive impairment caused by ingestible alcohol
	4009705	Age-related cognitive decline
	4022572	Disturbance of cognitive learning

Condition	Concept ID	Concept name
	4023989	Cognitive perceptual pattern
	4047110	Language-related cognitive disorder
	4297400	Mild cognitive disorder
	4333671	Age-associated memory impairment
	40480615	Cognitive disorder
	40482301	Residual cognitive deficit as late effect of cerebrovascular accident
	42535016	Cognitive deficit in communication skills
	42535017	Cognitive deficit in visuospatial function
	42535018	Cognitive deficit in psychomotor function
	42535681	Cognitive deficit due to and following ischaemic cerebrovascular accident
	42535682	Cognitive deficit due to and following hemorrhagic cerebrovascular accident
	42535706	Cognitive deficit due to and following embolic cerebrovascular accident
	42537139	Dissociative neurological symptom disorder co-occurrent with cognitive symptoms
	42539256	Cognitive deficit due to and following cerebrovascular disease
	42539270	Cognitive deficit due to and following nontraumatic subarachnoid hemorrhage
	42539271	Cognitive deficit due to and following nontraumatic intracerebral hemorrhage
	45765899	Moderate cognitive impairment
	45765900	Severe cognitive impairment
	46271045	Neurocognitive disorder
	40480615	Cognitive disorder
	4182210	Dementia
	42690615	Difficulty remembering past events
	42689981	Difficulty remembering people
	42689830	Difficulty remembering places
	42689982	Difficulty remembering routines
	42690112	Does not remember past events
	42690742	Does not remember people
	42690113	Does not remember places
	42689849	Does not remember routines
	443432	Impaired cognition

Condition	Concept ID	Concept name
	4103572	Organic memory impairment
	4135668	Poor auditory sequential memory
	4085496	Poor long-term memory
	4084412	Poor short-term memory
	4131380	Poor visual sequential memory
	42690368	Unable to remember past events
	42690369	Unable to remember people
	42690647	Unable to remember places
	42690370	Unable to remember routines
	4141586	Uncompensated short term memory deficit
	4043378	Frontotemporal dementia
Dyspnoea	4144682	Expiratory Dyspnoea
	4192279	Medical Research Council Dyspnoea scale grade 5
	4193263	Medical Research Council Dyspnoea scale grade 2
	4206307	Paroxysmal nocturnal Dyspnoea
	4212233	Dyspnoea after eating
	4217021	Dyspnoea, class II
	4219335	Dyspnoea, class III
	4219740	Borg Breathlessness Score finding
	4228754	Dyspnoea associated with AIDS
	4244276	Paroxysmal Dyspnoea
	4248284	Dyspnoea, class IV
	4263848	Dyspnoea on exertion
	4307188	Medical Research Council Dyspnoea scale grade 3
	4310172	Medical Research Council Dyspnoea scale grade 4
	35610140	eMRC (extended Medical Research Council) dyspnoea scale grade 2
	35610141	eMRC (extended Medical Research Council) dyspnoea scale grade 3
	35610142	eMRC (extended Medical Research Council) dyspnoea scale grade 5a
	35610143	eMRC (extended Medical Research Council) dyspnoea scale grade 4
	35610144	eMRC (extended Medical Research Council) dyspnoea scale grade 5b
	36685569	mMRC (modified Medical Research Council) dyspnoea scale grade 2

Condition	Concept ID	Concept name
	36685570	mMRC (modified Medical Research Council) dyspnoea scale grade 3
	36685571	mMRC (modified Medical Research Council) dyspnoea scale grade 4
	312437	Dyspnoea
Sleep disturbance	435657	Dyssomnia
	4204989	Disturbance in sleep behavior
	40480927	Sleep dysfunction with sleep stage disturbance
	40482260	Sleep dysfunction with arousal disturbance
	4132137	Sleep pattern disturbance
	435524	Sleep disorder
	42690122	Does not sleep
	374905	Non-organic sleep disorder
	443544	Organic sleep disorder
	4200883	Poor sleep pattern
	4215402	Primary insomnia
	37110488	Chronic insomnia
	4102985	Nonorganic insomnia
	4086851	Cannot sleep at all
	4115402	Difficulty sleeping
	42689992	Difficulty sleeping without sedation
	42689991	Difficulty sleeping with sedation
	43530738	Disruptions of 24 hour sleep-wake cycle
	4204989	Disturbance in sleep behavior
	42690122	Does not sleep
	42690715	Does not sleep without sedation
	42690123	Does not sleep with sedation
	434172	Insomnia with sleep apnea
	436522	Irregular sleep-wake pattern
	3173994	Poor sleep hygiene
	4305303	Sleep deprivation
	42690380	Unable to sleep without sedation
	42690379	Unable to sleep with sedation
Anaemia and haematinic deficiency	434622	deficiency anaemias
	4306430	Ferritin level low
	37016158	Low serum ferritin

Condition	Concept ID	Concept name
	136949	Refractory anaemia with excess blasts (clinical)
	137829	Aplastic anaemia
	140065	Pure red cell aplasia
	140681	Constitutional aplastic anaemia
	432282	Sideroblastic anaemia
	432295	Pernicious anaemia
	432588	Megaloblastic anaemia due to vitamin B-12 deficiency
	432875	Anaemia due to chronic blood loss
	433168	Iron deficiency anaemia secondary to inadequate dietary iron intake
	434894	Acute posthemorrhagic anaemia
	435503	Hemolytic anaemia
	435789	Megaloblastic anaemia
	436659	Iron deficiency anaemia
	437247	Anaemia of chronic disease
	437834	Non-autoimmune hemolytic anaemia
	438722	Non megaloblastic anaemia associated with nutritional deficiency
	439777	Anaemia
	440977	Megaloblastic anaemia due to folate deficiency
	440979	Acquired hemolytic anaemia
	441258	Anaemia in neoplastic disease
	441269	Autoimmune hemolytic anaemia
	443961	Anaemia of chronic renal failure
	602143	Anaemia due to chronic kidney disease stage 1
	603007	Pernicious anaemia due to autoimmune disorder
	603011	Vitamin B12 deficiency anaemia following total gastrectomy
	605330	Restless leg syndrome due to iron deficiency anaemia
	606065	Congenital megaloblastic anaemia due to transcobalamin II deficiency
	606921	Post gastrectomy iron deficiency anaemia
	606928	Iron deficiency anaemia due to celiac disease
	606940	Drug-induced non autoimmune hemolytic anaemia
	607426	Vitamin B12 deficiency anaemia due to chronic atrophic gastritis
	608596	Anaemia caused by antineoplastic agent

Condition	Concept ID	Concept name
	3661626	Megaloblastic anaemia due to dihydrofolate reductase deficiency
	4002495	Refractory anaemia with excess blasts in transformation (clinical)
	4003185	Refractory anaemia (clinical)
	4003186	Refractory anaemia with ringed sideroblasts (clinical)
	4006467	Anaemia due to infection
	4006468	Anaemia due to physical agent
	4008273	Coombs negative hemolytic anaemia
	4008663	Megaloblastic anaemia due to exfoliative dermatitis
	4009306	Anaemia due to copper deficiency
	4009785	Anaemia due to membrane defect
	4015896	Anaemia due to niacin deficiency
	4019001	Regenerative anaemia
	4021911	Megaloblastic anaemia due to poor nutrition
	4031699	Humoral immunologic aplastic anaemia
	4032006	Dimorphic anaemia
	4032352	Hemolytic anaemia due to hyperbaric oxygen
	4034963	Megaloblastic anaemia, thiamine-responsive, with diabetes mellitus and sensorineural deafness
	4035974	Drug-induced enzyme deficiency anaemia
	4039536	Autoimmune hemolytic anaemia due to complement
	4044728	Thiamine-responsive megaloblastic anaemia
	4045142	Megaloblastic anaemia due to gastrectomy
	4046563	G-6-PD class I variant anaemia
	4071444	Anaemia of thyroid dysfunction
	4079181	G-6-PD class III variant anaemia
	4082253	Unstable hemoglobin disease
	4082917	Drug-induced immune hemolytic anaemia, immune complex type
	4085853	Hemolytic anaemia due to babesiosis
	4092893	Anaemia due to vitamin E deficiency
	4096927	Megaloblastic anaemia due to vitamin B-12 malabsorption with proteinuria
	4097961	Acute megaloblastic anaemia due to dialysis
	4098008	Folate deficiency anaemia due to malabsorption
	4098009	Folate deficiency anaemia due to liver disorders

Condition	Concept ID	Concept name
	4098013	Hemolytic anaemia due to hexokinase deficiency
	4098017	Primary cold-type hemolytic anaemia
	4098018	Mechanical hemolytic anaemia
	4098019	Toxic hemolytic anaemia
	4098027	Aplastic anaemia due to radiation
	4098131	Myelophthisic anaemia
	4098145	Idiopathic aplastic anaemia
	4098627	Idiopathic hypochromic anaemia
	4098740	Vitamin B12 deficiency anaemia due to malabsorption with proteinuria
	4098746	Hemolytic anaemia due to pyruvate kinase deficiency
	4098747	Anaemia due to disorders of nucleotide metabolism
	4098760	Transient hypoplastic anaemia
	4098762	Pyridoxine-responsive sideroblastic anaemia
	4099508	Refractory anaemia without sideroblasts, so stated
	4099603	Megaloblastic anaemia due to hemodialysis
	4100962	Megaloblastic anaemia due to impaired absorption of folate
	4100985	Iron deficiency anaemia due to dietary causes
	4100987	Folate deficiency anaemia, drug-induced
	4100991	Hemolytic anaemia due to triose phosphate isomerase deficiency
	4100998	Aplastic anaemia due to toxic cause
	4101000	Secondary sideroblastic anaemia due to drugs and toxins
	4101001	Chronic anaemia
	4101458	Combined B12 and folate deficiency anaemia
	4101573	Vitamin C deficiency anaemia
	4101582	Aplastic anaemia due to chronic disease
	4101583	Aplastic anaemia due to infection
	4101584	Secondary sideroblastic anaemia due to disease
	4104541	Anaemia due to pentose phosphate pathway defect
	4105643	Myasthenic syndrome due to pernicious anaemia
	4114026	Normocytic anaemia
	4116343	Normocytic anaemia following acute bleed
	4120446	Iron deficiency without anaemia
	4120448	Normocytic anaemia due to aplasia

Condition	Concept ID	Concept name
	4120450	Normocytic anaemia due to chronic blood loss
	4121106	Microcytic anaemia
	4121110	Selective malabsorption of cyanocobalamin
	4121115	Sickle cell anaemia with high hemoglobin F
	4122079	Deficiency anaemias, excluding iron
	4122923	Dilutional anaemia
	4122924	Anaemia of renal disease
	4122927	Combined deficiency anaemia
	4125491	Megaloblastic anaemia due to dietary causes
	4125493	Vegan's anaemia
	4125630	Chronic non-spherocytic hemolytic anaemia
	4130191	Secondary warm autoimmune hemolytic anaemia
	4130680	Autoimmune hemolytic anaemia, categorised by antibody class AND/OR complement
	4131127	Secondary autoimmune hemolytic anaemia
	4131914	Sickle cell anaemia with coexistent alpha-thalassemia
	4131915	Primary (idiopathic) autoimmune hemolytic anaemia
	4131917	Paroxysmal cold hemoglobinuria
	4131919	Drug-induced immune hemolytic anaemia, hapten type
	4132085	Anaemia due to alloimmune destruction of transfused red cells
	4135931	Anaemia of endocrine disorder
	4143167	Autoimmune hemolytic anaemia due to IgA plus complement
	4143351	Folate deficiency anaemia due to dietary causes
	4143629	Anaemia due to mechanical damage
	4144077	G-6-PD class II variant anaemia
	4144811	Anaemia due to zinc deficiency
	4145277	HNSHA due to pyrimidine-5'-nucleotidase deficiency
	4146086	Constitutional aplastic anaemia with malformation
	4146088	Aplastic anaemia due to drugs
	4146771	Anaemia in ovarian carcinoma
	4146936	Drug-induced autoimmune hemolytic anaemia
	4147365	Anaemia of adrenal dysfunction
	4147491	Vitamin B12 deficiency anaemia due to dietary causes
	4147600	Megaloblastic anaemia due to pancreatic insufficiency

Condition	Concept ID	Concept name
	4147911	Megaloblastic anaemia due to inborn errors of metabolism
	4148471	Fanconi's anaemia
	4149183	Megaloblastic anaemia due to error of folate metabolism
	4150499	Refractory megaloblastic anaemia
	4150547	Anaemia secondary to renal failure
	4151502	Hemolytic anaemia due to Clostridium welchii
	4155187	Traumatic cardiac hemolytic anaemia
	4156842	Intracorpuscular hemolytic anaemia
	4157495	Sideropenic anaemia with reticuloendothelial siderosis
	4158891	Anaemia due to isoimmunization
	4159651	Traumatic hemolytic anaemia
	4159748	Hand-foot syndrome in sickle cell anaemia
	4160238	Simple chronic anaemia
	4160887	Cold autoimmune hemolytic anaemia
	4168286	Anaemia due to starvation
	4168772	Megaloblastic anaemia due to chronic hemolytic anaemia
	4171026	HNSHA due to NADH diaphorase deficiency
	4172446	Microcytic normochromic anaemia
	4173028	Idiopathic sideroblastic anaemia
	4174412	Anaemia due to copper
	4176884	Hemolytic anaemia due to nonlymphoid neoplasm
	4177177	Cellular immunologic aplastic anaemia
	4178677	Congenital nonspherocytic hemolytic anaemia due to inborn error of metabolism
	4184200	Secondary aplastic anaemia
	4184603	Megaloblastic anaemia due to Zollinger-Ellison syndrome
	4184758	Acquired aplastic anaemia
	4195171	Normocytic hypochromic anaemia
	4195271	Megaloblastic anaemia due to vegetarianism
	4201444	Anaemia due to riboflavin deficiency
	4203291	HNSHA due to increased adenosine deaminase activity
	4207240	Anaemia due to intrinsic red cell abnormality

Condition	Concept ID	Concept name
	4211348	Aplastic anaemia associated with pancreatitis
	4211695	Acute pure red cell aplasia
	4213893	Achlorhydric anaemia
	4214023	G-6-PD class V variant anaemia
	4215784	Autoimmune hemolytic anaemia due to IgM
	4215791	Acute megaloblastic anaemia secondary to total parenteral nutrition
	4216915	Hemoglobin S sickling disorder with crisis
	4217370	Aase syndrome
	4218100	Hemolytic anaemia due to drugs
	4218974	Hypoplastic anaemia
	4219253	Anaemia due to arsenic hydride
	4219359	G-6-PD class IV variant anaemia
	4219853	Warm autoimmune hemolytic anaemia
	4220697	Acute megaloblastic anaemia
	4221567	Megaloblastic anaemia due to disease of small intestine
	4223031	Anaemia associated with AIDS
	4223896	Mycoplasmal anaemia
	4225810	Aplastic anaemia associated with AIDS
	4228194	Congenital hypoplastic anaemia
	4228444	Acquired hemolytic anaemia associated with AIDS
	4231887	Secondary acquired sideroblastic anaemia
	4234973	Chronic acquired pure red cell aplasia
	4235788	Familial megaloblastic anaemia
	4238904	Autoimmune hemolytic anaemia due to IgA
	4241982	Congenital dyserythropoietic anaemia, type I
	4242111	HNSHA due to phosphoglycerate kinase deficiency
	4243831	Anaemia of pituitary deficiency
	4243950	Megaloblastic anaemia due to blind loop syndrome
	4244129	Anaemia due to decreased red cell production
	4246105	Hemolytic anaemia with emphysema AND cutis laxa
	4247416	Megaloblastic anaemia due to congenital deficiency of intrinsic factor
	4250028	Acute megaloblastic anaemia due to nitrous oxide
	4254249	HNSHA due to pyruvate kinase deficiency

Condition	Concept ID	Concept name
	4254380	Coombs positive hemolytic anaemia
	4258685	HNSHA due to triosephosphate isomerase deficiency
	4260689	Anaemia due to multiple mechanisms
	4261354	Megaloblastic anaemia due to decreased intake of vitamin B-12
	4262948	Microcytic hypochromic anaemia
	4263315	Normocytic normochromic anaemia
	4264046	Sickle cell-hemoglobin E disease
	4265915	HNSHA due to diphosphoglycerate mutase deficiency
	4268894	Acute megaloblastic anaemia due to severe illness
	4269764	Glucose-6-phosphate dehydrogenase deficiency anaemia
	4269919	Autoimmune hemolytic anaemia due to IgG plus complement
	4271197	Idiopathic paroxysmal cold hemoglobinuria
	4278920	Anaemia due to lead
	4280070	Antibody-mediated anaemia
	4280354	Nutritional anaemia
	4282785	Megaloblastic anaemia due to nontropical sprue
	4284415	Megaloblastic anaemia due to increased requirements
	4286660	Congenital dyserythropoietic anaemia, type II
	4287402	Anaemia of parathyroid dysfunction
	4287574	Megaloblastic anaemia due to error of cobalamin metabolism
	4291002	Megaloblastic anaemia due to drugs
	4297024	Hemolytic anaemia due to Bartonella
	4297537	Hemolytic anaemia due to infection
	4298690	Immunologic aplastic anaemia
	4298975	Hemolytic anaemia due to malaria
	4300295	Drug-induced sideroblastic anaemia
	4303199	Anaemia due to pantothenic deficiency
	4306199	Perinatal anaemia
	4307469	Sports anaemia
	4307799	Anaemia due to diabetes mellitus
	4308062	Diaphyseal dysplasia with anaemia
	4308125	Macrocytic anaemia

Condition	Concept ID	Concept name
	4311676	Anaemia due to vitamin A deficiency
	4312008	Anaemia due to substance
	4312853	Anaemia due to vitamin B-6 deficiency
	4313413	Anaemia due to chlorate
	4313581	Hapten type high affinity hemolytic anaemia
	4314111	Non megaloblastic anaemia due to alcoholism
	4318674	Chronic idiopathic autoimmune hemolytic anaemia
	4319914	Anaemia due to radiation
	4323223	Anaemia due to medication
	4329173	Anaemia of gonadal dysfunction
	4330322	Anaemia due to disturbance of proliferation AND/OR differentiation of hematopoietic stem cells
	4336555	G-6-PD variant enzyme deficiency anaemia
	4338370	Megaloblastic anaemia due to alcoholism
	4338976	Megaloblastic anaemia due to tropical sprue
	35624317	Hemolytic anaemia due to adenylate kinase deficiency
	35624756	Anaemia due to and following chemotherapy
	36680584	Autosomal dominant aplasia and myelodysplasia
	36713571	Megaloblastic anaemia due to vitamin B12 deficiency secondary to intestinal disease
	36713572	Vitamin B12 deficiency anaemia caused by drug
	36713573	Acquired iron deficiency anaemia due to increased iron requirement
	36713763	Autoimmune hemolytic anaemia mixed type
	36715009	Adult-onset autosomal recessive sideroblastic anaemia
	36715492	Megaloblastic anaemia due to folate deficiency due to increased requirement
	36715580	Acquired thiamine deficiency anaemia
	36715584	Refractory anaemia with ringed sideroblasts associated with marked thrombocytosis
	36716029	Hyperuricemia, anaemia, renal failure syndrome
	36716126	Iron-refractory iron deficiency anaemia
	36716259	Pancreatic insufficiency, dyserythropoietic anaemia, calvarial hyperostosis syndrome
	36716460	X-linked congenital dyserythropoietic anaemia with thrombocytopenia

Condition	Concept ID	Concept name
	37016121	Anaemia following acute postoperative blood loss
	37016151	Aplastic anaemia caused by antineoplastic agent
	37017132	Anaemia co-occurrent with human immunodeficiency virus infection
	37017165	GATA binding protein 1 related thrombocytopenia with dyserythropoiesis
	37017285	Acquired hemolytic anaemia co-occurrent with human immunodeficiency virus infection
	37018722	Anaemia caused by zidovudine
	37019055	Aplastic anaemia co-occurrent with human immunodeficiency virus infection
	37019193	Anaemia co-occurrent and due to chronic kidney disease stage 3
	37110070	Mitochondrial myopathy with sideroblastic anaemia syndrome
	37110336	Acquired iron deficiency anaemia due to decreased absorption
	37110727	Nonspherocytic hemolytic anaemia due to deficiency of adenosinetriphosphatase
	37110923	Severe congenital hypochromic anaemia with ringed sideroblasts
	37111627	Central nervous system calcification, deafness, tubular acidosis, anaemia syndrome
	37116297	Secondary autoimmune hemolytic anaemia co-occurrent and due to chronic inflammatory disease
	37116298	Secondary autoimmune hemolytic anaemia co-occurrent and due to lymphoproliferative disorder
	37116300	Secondary autoimmune hemolytic anaemia co-occurrent and due to rheumatic disorder
	37116301	Secondary autoimmune hemolytic anaemia co-occurrent and due to ulcerative colitis
	37117740	Secondary autoimmune hemolytic anaemia co-occurrent and due to systemic lupus erythematosus
	37119138	Iron deficiency anaemia due to blood loss
	37204236	X-linked dyserythropoietic anaemia with abnormal platelets and neutropenia
	37204287	Hemoglobinopathy Toms River
	37204551	Hereditary isolated aplastic anaemia
	37312032	Anaemia due to chronic infectious disease
	37395652	Anaemia in chronic kidney disease stage 5
	37397036	Autosomal recessive sideroblastic anaemia

Condition	Concept ID	Concept name
	37398911	Anaemia in chronic kidney disease stage 4
	40478891	Erythropoietin resistance in anaemia of chronic kidney disease
	40599994	X chromosome-linked sideroblastic anaemia
	42536530	Hereditary vitamin B12 deficiency anaemia
	42536531	Hereditary folate deficiency anaemia
	42537687	Anaemia due to metabolic disorder
	42872405	Anaemia, pre-end stage renal disease on erythropoietin protocol
	44783626	Pulmonary arterial hypertension associated with chronic hemolytic anaemia
	44806268	Refractory anaemia with multilineage dysplasia
	44810002	Recurrent anaemia
	45768812	Anaemia in chronic kidney disease
	45768813	Anaemia in end stage renal disease
	45768941	Chronic hemolytic anaemia
	45773534	Anaemia in malignant neoplastic disease
	46272744	Hypochromic microcytic anaemia with iron overload
Hypertension	312648	Benign essential hypertension
	4215640	Benign essential hypertension complicating AND/OR reason for care during childbirth
	4034031	Benign essential hypertension complicating AND/OR reason for care during pregnancy
	4148205	Benign essential hypertension complicating AND/OR reason for care during puerperium
	4269358	Benign essential hypertension in obstetric context
	320128	Essential hypertension
	4083723	Essential hypertension complicating AND/OR reason for care during childbirth
	4302591	Essential hypertension complicating AND/OR reason for care during pregnancy
	4321603	Essential hypertension complicating AND/OR reason for care during puerperium
	4217486	Essential hypertension in obstetric context
	4058987	High-renin essential hypertension
	4159755	Labile essential hypertension
	4263067	Low-renin essential hypertension
	317898	Malignant essential hypertension

Condition	Concept ID	Concept name
	45757787	Postpartum pre-existing essential hypertension
	4180283	Systolic essential hypertension
Diabetes	201826	Type 2 diabetes mellitus
	201254	Type 1 diabetes mellitus
	443731	Renal disorder due to type 2 diabetes mellitus
	200687	Renal disorder due to type 1 diabetes mellitus
	443729	Peripheral circulatory disorder due to type 2 diabetes mellitus
	318712	Peripheral circulatory disorder due to type 1 diabetes mellitus
	376065	Disorder of nervous system due to type 2 diabetes mellitus
	377821	Disorder of nervous system due to type 1 diabetes mellitus
	443733	Disorder of eye due to type 2 diabetes mellitus
	42538169	Disorder of eye due to type 1 diabetes mellitus
	443732	Disorder due to type 2 diabetes mellitus
	435216	Disorder due to type 1 diabetes mellitus
Osteoporosis	80502	Osteoporosis
	37204244	X-linked osteoporosis with fractures
	4109181	Osteoporosis with pseudoglioma
	44783850	Osteoporosis circumscripta
	36716194	Osteoporosis and oculocutaneous hypopigmentation syndrome
Chronic kidney disease	46271022	Chronic kidney disease
	75865	Disorder of the urinary system
	197331	Disorder of urinary tract
Skin Ulcer	4262920	Skin ulcer
	46269755	Chronic non-pressure ulcer of calf extending to fat level
	46269752	Chronic non-pressure ulcer of ankle extending to fat level
Ischeamic heart disease	4185932	Ischeamic heart disease
Heart Failure	316139	Heart failure
Cerebrovascular disease	381591	Cerebrovascular disease
Peripheral vascular disease	321052	Peripheral vascular disease
Atrial fibrilation	313217	Atrial fibrilation
Heart valve disease	4281749	Heart valve disorder

Condition	Concept ID	Concept name
Hypotension/syncope	317002, 40350983	Low blood pressure
	40316030	Hypotension
	319041	Orthostatic hypotension
	135360, 40498271	Syncope
Foot problem	4101512	Foot problem
	4268887	Chiropody follow-up
	4053100	Domiciliary chiropody
	4136647	Seen by community-based podiatrist
	4138349	Seen by community-based podiatry service
	4140790	Seen by hospital-based podiatrist
	4140924	Seen by hospital-based podiatry service
	42539590	Seen by podiatric surgeon
	4083436	Seen by podiatrist
	4139895	Seen by podiatry service
	4085778	Seen in chiropody clinic
	4139217	Under care of community-based podiatrist
	4139218	Under care of hospital-based podiatrist
	42539494	Under care of podiatric surgeon
	4139705	Under care of podiatrist
	4067069	Callosity
Arthritis	4291025	Arthritis
	40555828	Arthritis
Chronic Respiratory disease	4063381	Chronic disease of the respiratory system
	261325	Pulmonary emphysema
	255573	Chronic obstructive lung disease
	317009	Asthma
Peptic ulcer	4027663	Peptic ulcer
	4057060	Acute Peptic ulcer
	4134146	Chronic Petic ulcer
Thyroid disease	141253	Disorder of thyroid gland
	4194160	Thyroid function tests abnormal
Fragility fracture	44791986	Fragility fracture
	40480160	Osteoporotic fracture

Condition	Concept ID	Concept name
	4001142	Osteopathies, chondropathies and acquired musculoskeletal deformities
	44791986	Fragility fracture
	4174520	Fracture of vertebral column
	4050747	Fracture of upper limb
	4053828	Fracture of thoracic spine
	4302740	Fracture of sternum
	4142905	Fracture of rib
	4013613	Fracture of lumbar spine and/or pelvis
	4278672	Fracture of forearm
	442560	Fracture of femur
	4129393	Fracture of cervical spine
	4015350	Fracture at wrist and/or hand level
	4001458	Fatigue fracture of vertebra
	4344386	Disorder of continuity of bone
	4222001	Collapse of vertebra
Urinary System disease	4127562	Infective cystitis
	81902	Urinary tract infectious disease
	198199	Pyelonephritis
	195862	Urethritis
	4183440	Subacute cystitis
	36715430	Sepsis due to urinary tract infection
	4159655	Infection of bladder catheter
	201338	Urethral fistula
	4284706	Urethrotrigonitis
	196464	Rapidly progressive glomerulonephritis
	4054992	Rapidly progressive nephritic syndrome
	4127562	Infective Cystitis
	194081	Acute cystitis
	4189531	Acute nephritis
	4286024	Acute pyelonephritis
	4056023	Acute pyonephrosis
	19351000	Acute glomerulonephritis
	4284564	Acute infectious tubulointerstitial nephritis
	4126301	Idiopathic crescentic glomerulonephritis
Dizziness	4223938	Dizziness

Condition	Concept ID	Concept name
	433316	Dizziness and giddiness
	43021417	Dizziness due to drug
	42539141	Dizziness following neck extension
	45769954	Dizziness on lying still
	44806935	Dizziness on neck extension
	4250121	Dizziness on standing up
	1333255	Dizziness or light-headedness
	4012520	Dizziness present
	4198449	Dizzy spells
	1340313	Exacerbation of dizziness
	4012691	Exertional dizziness
	905035	Fainting or dizziness
	4297376	Lightheadedness
	4337455	Multisensory dizziness
	4097171	Oscillation of surroundings
	4011333	Persistent postural perceptual dizziness
	4012243	Postural dizziness
	40768436	Things occurring with dizziness, loss of balance or spinning sensation [PhenX]
Parkinsonism and tremor	381270	Parkinson's disease
	36716783	Atypical Parkinsonism
	37110549	Functional parkinsonism
	4140090	Parkinsonism
	372604	Movement disorder
	443782	Tremor

ANNEX V: ENCePP checklist for study protocols

Doc.Ref. EMA/540136/2009

ENCePP Checklist for Study Protocols (Revision 4)

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

Study title:

P4-C2-013 and P4-C2-014
DARWIN EU® - Determinants for use of GLP1 receptor agonists – a drug utilisation study

EU PAS Register® number: NA

Study reference number (if applicable): *This is a routinely repeated study from P3-C1-008 with EUPAS1000000223.*

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]	☒	☐	☐	9.3
	1.1.2 End of data collection[2]	☒	☐	☐	
	1.1.3 Progress report(s)	☐	☐	☒	6
	1.1.4 Interim report(s)	☐	☐	☒	
	1.1.5 Registration in the EU PAS Register®	☒	☐	☐	
	1.1.6 Final report of study results.	☒	☐	☐	

Comments:

Section 2: Research question		**Yes**	**No**	**N/A**	**Section Number**
2.1	Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	
	2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	8
	2.1.2 The objective(s) of the study?	☒	☐	☐	
	2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	9.5

[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 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

Section 3: Study design		**Yes**	**No**	**N/A**	**Section Number**
3.1	Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	9.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	9.2
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	9.6, 9.8
3.4	Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☒	
3.5	Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

Comments:

Section 4: Source and study populations		**Yes**	**No**	**N/A**	**Section Number**
4.1	Is the source population described?	☒	☐	☐	
4.2	Is the planned study population defined in terms of:				
	4.2.1 Study time period	☒	☐	☐	9.3
	4.2.2 Age and sex	☒	☐	☐	9.6
	4.2.3 Country of origin	☒	☐	☐	9.2
	4.2.4 Disease/indication	☒	☐	☐	9.6
	4.2.5 Duration of follow-up	☒	☐	☐	9.4
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	9.5

Comments:

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	9.6.1
5.2	Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☒	☐	☐	10
5.3	Is exposure categorised according to time windows?	☒	☐	☐	9.6.3
5.4	Is intensity of exposure addressed? (e.g. dose, duration)	☒	☐	☐	9.6.3
5.5	Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☒	☒	
5.6	Is (are) (an) appropriate comparator(s) identified?	☐	☐	☒	

Comments:

Section 6: Outcome definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
6.1	Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☐	☐	☒	
6.2	Does the protocol describe how the outcomes are defined and measured?	☐	☐	☒	
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)	☐	☒	☐	10

Comments:

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

Comments:

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

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☐	☐	☒	
9.3.3 Covariates and other characteristics?	☒	☐	☐	Annex II
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☒	

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	9.8.3
10.2 Is study size and/or statistical precision estimated?	☐	☐	☒	
10.3 Are descriptive analyses included?	☒	☐	☐	9.8
10.4 Are stratified analyses included?	☒	☐	☐	9.6.1, 9.8
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	
10.7 Does the plan describe methods for handling missing data?	☒	☐	☐	9.9
10.8 Are relevant sensitivity analyses described?	☐	☐	☒	

Comments:

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

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?	☒	☐	☐	10
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☐	☐	☒	
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	9.7

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?	☒	☐	☐	Annex III
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	☒	
13.3 Have data protection requirements been described?	☒	☐	☐	Annex III

Comments:

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	5

Comments:

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

Comments:

ANNEX VI: Glossary

Additional definitions are available in the EMA Glossary of terms https://www.ema.europa.eu/en/about-us/glossaries.

Aggregated Data

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

Benefit-Risk Assessment

Evaluation of the positive therapeutic effects of a medicine compared to its risks (e.g., side effects).

Common Data Model (CDM)

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

Complex Studies (C3)

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

Coordination Centre (CC)

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

Data Access

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

Data Quality Framework

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

Data Source

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

DARWIN EU®

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

EMA (European Medicines Agency)

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

Evidence Generation

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

Federated Network

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

GDPR (General Data Protection Regulation)

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

Health Technology Assessment (HTA)

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

Metadata

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

Off-the-Shelf Studies (OTS)

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

OHDSI (Observational Health Data Sciences and Informatics)

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

Patient-Level Data

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

OMOP (Observational Medical Outcomes Partnership)

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

Real-World Data (RWD)

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

Real-World Evidence (RWE)

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

Regulatory Decision-Making

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

Routine Repeated Studies (RR)

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

Study Protocol

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

Very Complex Studies (C4)

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

Name of the main author of the protocol: Marta Pineda Moncusi

Date: 17/October/2025

Signature: MPM