



Study Protocol

P5-C1-008/009/010

DARWIN EU® - Sex-specific incidence of selected adverse events for major cardio-metabolic drug classes

24/08/2026

Version 4.0

Authors: Xintong Li, Albert Prats-Urbe

Public

CONTENTS

LIST OF ABBREVIATIONS.....	5
1. TITLE	7
2. DESCRIPTION OF THE STUDY TEAM	7
3. ABSTRACT	8
4. AMENDMENTS AND UPDATES	10
5. MILESTONES	10
6. RATIONALE AND BACKGROUND	10
7. RESEARCH QUESTIONS AND OBJECTIVES.....	11
8. RESEARCH METHODS.....	11
8.1. Study design	11
Figure 1. Graphical depiction of the study design, statin as an example.	12
8.2. Follow-up	12
8.3. Study population with inclusion and exclusion criteria.....	13
8.4. Study setting and data sources	13
Table 1. Data sources.....	13
8.5. Study period	14
8.6. Variables	14
8.6.1. Exposure	14
8.6.2. Outcome	15
8.6.3. Covariates.....	15
Table 2. Co-mediations of interest for characteristics for each of the study treatment initiator cohorts.....	16
Table 3. Conditions/procedures that are potential risk factors to study outcomes.	18
8.7. Study size	19
8.8. Analysis	19
8.8.1. Federated network analyses	19
8.8.2. Data privacy protection	19
8.8.3. Statistical model specification and assumptions of the analytical approach considered	19
Objective 1.....	19
Objective 2.....	20
8.8.4. Output	20
8.9. Evidence synthesis.....	21
9. STRENGTHS AND LIMITATIONS	21
10. REFERENCES	21
11. ANNEXES.....	23
ANNEX I. Description of data sources.....	23
Belgium: IQVIA LPD Belgium (IQVIA Longitudinal Patient Database Belgium)	23
Finland: Finnish Care Register for Health Care (FinOMOP-THL).....	24
Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)	25
Spain: The Information System for Research in Primary Care (SIDIAP)	26
The Netherlands: Integrated Primary Care Information (IPCI).....	27
The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)	29
ANNEX II. Fitness for use assessment.....	31

ANNEX III. Operational and reporting considerations.....	34
ANNEX IV. Preliminary code list for exposures and outcomes.....	36
Table S1. Preliminary list of conditions definitions.	36
Table S2. Preliminary list of medicines definitions.....	36
ANNEX V. ENCePP checklist for study protocols	38
ANNEX VI. Mock tables and figures.....	44
MOCK Table 1. Baseline characteristics of lipophilic statins initiators, by data source.	44
MOCK Figures 1–8. Sex-specific incidence rates of study outcomes among lipophilic statins initiators, overall age group, by data source.	47
MOCK Figures 9–16. Sex-specific cumulative incidence of study outcomes among lipophilic statins initiators.....	47
MOCK Supplementary Table. Incidence rates per 100,000 person-years and 95% CI of study outcomes among on-treatment lipophilic statins initiator cohort.....	48

Study title	DARWIN EU® - Sex-specific incidence of selected adverse events for major cardio-metabolic drug classes
Protocol version	V4.0
Date	24/08/2026
EUPAS number	EUPAS1000001081
Active Substance	HMG CoA reductase inhibitors (statins) (ATC C10AA) GLP-1 receptor agonists (ATC A10BJ) SGLT2 inhibitors (ATC A10BK)
Medicinal Product	NA.
Research questions and objectives	The aim of this study is to estimate the sex-specific incidence of various adverse outcomes among males and females who initiate selected treatments used for cardiovascular diseases risk reduction. The objectives of this study are to: 1. Identify first-time treatment initiators of any drug in the class and describe the characteristics at the time (age, potential indication for index drug, comorbidities, co-medication), stratified by sex 2. Estimate the incidence of each adverse outcomes post-first initiation for any drug in the class, stratified by sex and age group.
Countries of study	Belgium, Finland, Germany, Spain, the Netherlands, the United Kingdom
Authors	Xintong Li (x.li@darwin-eu.org) Albert Prats-Urbe (a.prats-uribe@darwin-eu.org)

LIST OF ABBREVIATIONS

Acronyms/terms	Description
ATC	Anatomical Therapeutic Chemical
ACE inhibitors	Angiotensin-Converting Enzyme inhibitors
ARBs	Angiotensin II Receptor Blockers
BIFAP	Pharmacoepidemiological Research Database for Public Health Systems
CDM	Common Data Model
CKD	Chronic Kidney Disease
COPD	Chronic obstructive pulmonary disease
CPRD Aurum	Clinical Practice Research Datalink Aurum
CPRD GOLD	Clinical Practice Research Datalink GOLD
CVD	Cardiovascular Disease
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DDD	Defined Daily Dose
DPP-4	Dipeptidyl Peptidase-4
DRE	Digital Research Environment
DTZ	Data Transfer Zone
EHR	Electronic Health Records
EMA	European Medicines Agency
EMC	Erasmus University Medical Center
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
FinOMOP-THL	Finnish Care Register for Health Care
GERD	Gastroesophageal reflux disease
GLP-1	Glucagon-Like Peptide-1
GLP-1RA	Glucagon-like peptide-1 receptor agonists
GP	General Practitioner
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage
HMG-CoA	3-hydroxy-3-methylglutaryl coenzyme A
HUVM	Hospital Universitario Virgen Macarena
ICD	International Classification of Diseases
IDIAPJGol	Institut d'Investigació en Atenció Primària de Salut Jordi Gol
InGef RDB	InGef Research Database
IPCI	Integrated Primary Care Information
IQVIA DA Germany	IQVIA Disease Analyzer Germany
IQVIA LPD Belgium	IQVIA Longitudinal Patient Database Belgium
IRB	Institutional Review Board

Acronyms/terms	Description
LDH	Lithuanian Digital Health Platform
NAJS	Croatian National Public Health Information System
OMOP	Observational Medical Outcomes Partnership
OHDSI	Observational Health Data Sciences and Informatics
PCSK9	Proprotein Convertase Subtilisin/Kexin Type 9
PGH	Papageorgiou General Hospital
RAAS	Renin-angiotensin-aldosterone system
RxNorm	Medical prescription normalised
SD	Standard deviation
SGLT2	Sodium-Glucose Co-Transporter 2
SIDIAP	Information System for Research in Primary Care
SNOMED	Systemized Nomenclature of Medicine
T2DM	Type 2 Diabetes Mellitus
THL	Finnish Institute for Health and Welfare
ULSM-RT	Unidade Local de Saúde de Matosinhos Realtime Database
VID	Valencia Health System Integrated Dataset
WHO	World Health Organization

1. TITLE

DARWIN EU® - Sex-specific incidence of selected adverse events for major cardio-metabolic drug classes

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Xintong Li	University of Oxford
Data Scientist	Moronfoluwa Akintola	University of Oxford
Data Scientist	Edward Burn	University of Oxford
Clinical Domain Expert	Albert Prats Uribe	University of Oxford
Epidemiologist	Annika Jodicke	University of Oxford
Study Manager	Natasha Yefimenko	Erasmus Medical Center
Data source	Names	Data Partner Organisation*
IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)	Gargi Jadhav Hugo Vernooij Dina Vojinovic	IQVIA
Finnish Care Register for Health Care (FinOMOP-THL)	Toni Lehtonen Tuomo Susi Laura Salonen	Finnish Institute for Health and Welfare (THL)
IQVIA Disease Analyzer Germany (IQVIA DA Germany)	Gargi Jadhav Hugo Vernooij Dina Vojinovic	IQVIA
The Information System for Research in Primary Care (SIDIAP)	Elena Roel Herranz Talita Duarte Salles Laura Granés González Agustina Giuliodori Picco	Institut d'Investigació en Atenció Primària de Salut Jordi Gol (IDIAPJGol)
Integrated Primary Care Information (IPCI)	Katia Verhamme Mees Mosseveld	Erasmus University Medical Center (EMC)
Clinical Practice Research Datalink GOLD (CPRD GOLD)	Antonella Delmestri	University of Oxford - Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences

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

3. ABSTRACT

Title

DARWIN EU® - Sex-specific incidence of selected adverse events for major cardio-metabolic drug classes

Rationale and background

Although cardiovascular disease (CVD) is a leading cause of death in females, it is still often regarded as mainly affecting males. Females remain underdiagnosed, undertreated, and underrepresented in clinical trials, leading to treatment guidelines being based mainly on evidence from male populations. However, sex differences in drug responses can affect both treatment efficacy and the risk of adverse effects, highlighting the need for more sex-specific evidence.

Research questions and objectives

Research questions

Is the safety of statins, glucagon-like peptide-1 (GLP-1) receptor agonists, and Sodium-Glucose Co-Transporter 2 (SGLT2) inhibitors different between males and females? What baseline differences between male and female treatment initiators could explain any observed difference in adverse outcome incidence?

Objectives

The aim of this study is to estimate the sex-specific incidence of various adverse outcomes among males and females who initiate selected treatments used for cardiovascular diseases risk reduction.

The specific objectives of this study are to:

1. Identify first-time treatment initiators of any drug in the class and describe the characteristics at treatment initiation (age, potential indication for index drug, comorbidities, co-medication), stratified by sex.
2. Estimate the incidence of pre-specified adverse outcome post-first initiation for any drug in each pre-specified class, stratified by sex and age group.

Methods

Study design

A descriptive epidemiological new user cohort study. Individuals will be followed from the date of treatment initiation (index date) for up to 5 years, until death, exit from the data source, or the end of follow-up, whichever occurs first.

Population

The study will include all individuals who meet the eligibility criteria during the recruitment period (from 01/01/2014 to 31/12/2025). Eligibility criteria are:

- Initiation of one of the three classes of the medications of interest.
- Aged ≥ 18 years at the index date.
- At least 365 days of data source history prior to index date.
- No prior use of the medication from the index class any time prior to the index date.
- No outcomes of interest for the index class in the 365 days prior to the index date.

Variables

Exposure (drug classes of interest):

- 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors (statins)
- Glucagon-like peptide-1 (GLP-1) receptor agonists
- Sodium-glucose co-transporter 2 (SGLT2) inhibitors

Outcome:

For each of the exposure drug classes, specific outcomes will be measured:

- Statins: Muscle symptoms (myalgia, myopathy, rhabdomyolysis); New-onset type 2 Diabetes Mellitus
- GLP-1 receptor agonists (GLP-1RA): Gastro-intestinal adverse events (nausea, vomiting, delayed gastric emptying up to overt gastroparesis); Gallbladder disease (cholelithiasis, cholecystitis); Pancreatitis; Sarcopenia
- SGLT2 inhibitors: Genito-urinary events; Euglycemic Diabetic Ketoacidosis

Relevant covariates:

Age, sex, index year, potential indication for index drug, comorbidities, co-medication

Data sources

1. Belgium: IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)
2. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
3. Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)
4. Spain: The Information System for Research in Primary Care (SIDIAP)
5. The Netherlands: Integrated Primary Care Information (IPCI)
6. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Statistical analysis

For each of the three medication classes, two cohorts will be created: an “ever-treated” cohort, in which individuals will be followed regardless of treatment discontinuation or switching, and an “on-treatment” cohort, in which follow-up will end at treatment discontinuation or switching (defined using a 90-day grace window, and allowing switching between (sub-)class members).

In objective 1, baseline characteristics at treatment initiation, including demographics, comorbidity, and co-medication use, will be described. Standardised mean differences of baseline characteristics will be calculated between females and males. Treatment duration and number of prescriptions during the first treatment episode and within 5 years of initiation will be summarised.

In objective 2, incidence rates per 100,000 person-years and 95% Poisson confidence intervals will be estimated in both cohorts, stratified by sex and age.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	August 2026
Creation of Analytical code	July–August 2026
Execution of Analytical Code on the data	August 2026
Draft Study Report	September 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 and deliverables review rounds.

6. RATIONALE AND BACKGROUND

Cardiovascular diseases (CVD) are often considered to predominantly affect men, despite being the leading cause of death among women worldwide.[1] It has been widely acknowledged that women are underdiagnosed and undertreated with respect to CVD, contributing to persistent disparities in cardiovascular care.[1,2]

Furthermore, women remain underrepresented in cardiovascular clinical trials, and treatment efficacy and safety are rarely reported separately by sex.[2,3] Consequently, although the need for greater consideration of sex differences in CVD has been increasingly recognised, clinical guidelines continue to extrapolate evidence derived predominantly from male-dominated clinical trials and, in most situations, do not provide women-specific recommendations.[2]

Nonetheless, women present important differences in both pharmacokinetics and pharmacodynamics, which may influence the effectiveness and safety of cardiovascular therapies. Pharmacokinetic differences, including variations in body composition, hepatic clearance, renal excretion, and interactions with hormonal contraceptives, together with pharmacodynamic differences, such as changes in receptor expression and receptor affinity, may lead to differences in drug efficacy and the risk of adverse events.[3]

As part of the broader European Medicines Agency's (EMA) Women's Health Initiative, the current study aims to complement the limited evidence by comparing the incidence of various adverse outcomes among women and men initiating selected treatments used for CVD risk reduction. Due to the limited availability of gender information within the DARWIN EU® Network, the current study will focus on sex, namely male and female, instead of gender. The findings are expected to contribute to the evidence base on potential sex-related differences in the safety of cardiovascular pharmacotherapy and support more informed clinical and regulatory decision-making.

7. RESEARCH QUESTIONS AND OBJECTIVES

Research questions

Is the incidence of selected adverse events among statin, glucagon-like peptide-1 (GLP-1) receptor agonist (GLP-1RA), and sodium-glucose co-transporter 2 (SGLT2) inhibitor initiators different between males and females? What baseline differences between male and female treatment initiators could explain any observed difference in adverse outcomes incidence?

Research objectives

The aim of this study is to estimate the sex-specific incidence of various adverse outcomes among males and females who initiate selected treatments used for cardiovascular diseases risk reduction.

The specific objectives of this study are to:

1. Identify first-time treatment initiators of any drug in the class and describe the characteristics at the time (age, potential indication for index drug, comorbidities, co-medication), stratified by sex.
2. Estimate the incidence of each adverse outcome post-first initiation for any drug in the class, stratified by sex and age group.

8. RESEARCH METHODS

8.1. Study design

Cohort studies will be conducted using routinely collected health data from 6 data sources from 6 countries across Europe. The study will comprise:

- A patient-level characterisation, which will be conducted to address objective 1, assessing the characteristics of individuals who initiate the medications of interest.
- A new user cohort study, which will be conducted to address objective 2, assessing the incidence rates and cumulative incidence of outcomes of interests among individuals who initiate the medications of interest, depicted in [Figure 1](#).

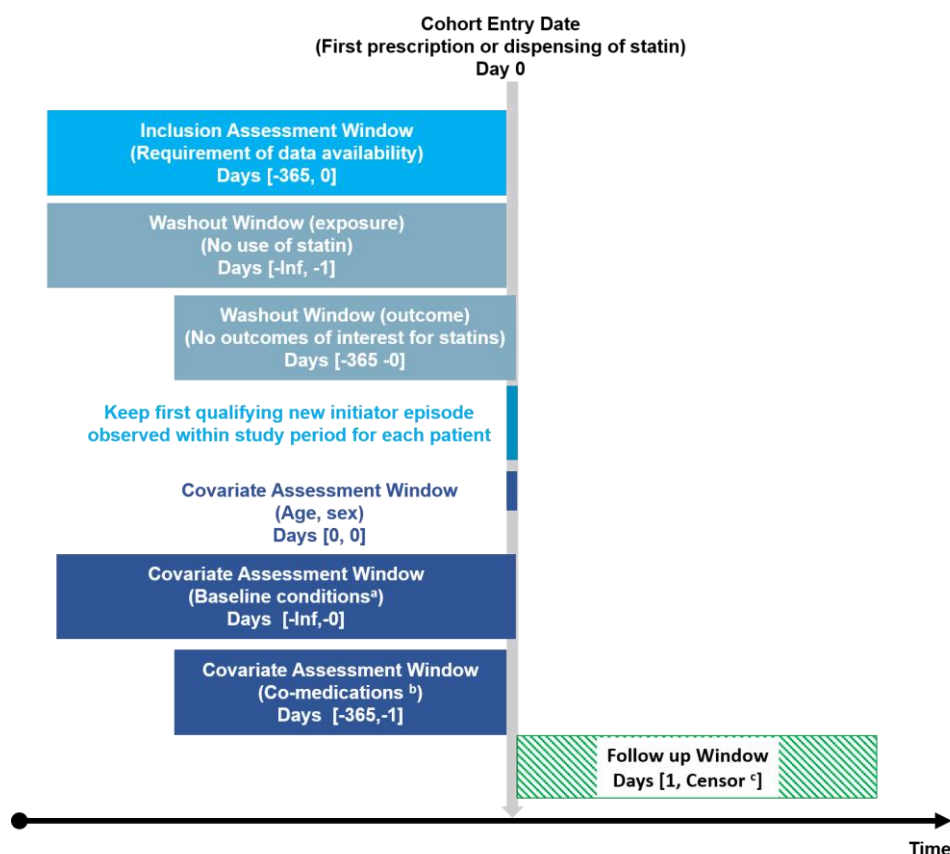


Figure 1. Graphical depiction of the study design, statin as an example.

- Baseline conditions include comorbidities, potential indications for study medications, and potential risk factors for study outcomes.
- Co-mediations of interest specific to drug class.
- Earliest of: outcome of interest (e.g. muscle symptoms or new-onset type 2 diabetes mellitus), switching or discontinuation of study drugs (only for the on-treatment cohort, based on drug subclass), death, disenrollment, 5 years of follow-up, end of the study period.

8.2. Follow-up

For the new user cohort study, follow-up will start on the date of first prescription or dispensation of each medication class.

The follow-up will be defined in two different ways: “ever-treated” and “on-treatment”.

For the “ever-treated” cohort, follow-up will end on the earliest of: loss to follow-up, death, 5 years since initiation, end of observation period (the latest available data), or the study end date of 31/12/2025.

For the “on-treatment” cohort, follow-up will end on the earliest of: loss to follow-up, death, end of treatment episode (discontinuation defined using 90-day grace window, or switching, based on medication subclasses), 5 years since initiation, end of observation period (the latest available data), or the study end date of 31/12/2025.

8.3. Study population with inclusion and exclusion criteria

All objectives:

Inclusion criteria

- Present in the period of 01/01/2014 to 31/12/2025 (or the latest available date)
- Minimum 365 days of available history before the index date
- Drug prescription or dispensing record of the three classes of medication during the study period
- Aged ≥ 18 years at the index date

Exclusion criteria

- Prescription or dispensing of any drug from the index class any time prior to the index date
- Having experienced the outcomes of interest (for the index class) before the index date
- Initiated more than one ingredient from the same drug class on the index date

8.4. Study setting and data sources

This study will be conducted using routinely collected data from 6 data sources in the DARWIN EU[®] Network from 6 European countries. All data were *a priori* mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

Table 1. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals*	Calendar period covered by each data source	Contributing to
Belgium	IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)	Primary care	Outpatient General Practitioner (GP) Care	152,500	2016 – 03/2026	All objectives
Finland	Finnish Care Register for Health Care (FinOMOP-THL)	Primary care, secondary care	Registries	6,154,220	2015 – 12/2025	All objectives
Germany	IQVIA Disease Analyzer Germany (IQVIA DA Germany)	Primary care, specialist	Outpatient GP Care	4,715,400	1993 – 06/2025	All objectives
Spain	The Information System for Research in Primary Care (SIDIAP)	Primary care, secondary care	Outpatient GP Care	6,068,000	2006 – 12/2024	All objectives

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals*	Calendar period covered by each data source	Contributing to
The Netherlands	Integrated Primary Care Information (IPCI)	Primary care	Outpatient GP Care	1,339,300	2006 – 12/2025	All objectives
The United Kingdom	Clinical Practice Research Datalink GOLD (CPRD GOLD)	Primary care	Outpatient GP Care	2,371,680	1989 – 11/2025	All objectives

*Number of active individuals: Active Patients is defined as the maximum number of individuals in an observation period, in the last 6 months of data for that source, accessed on 06/07/2026.

Data sources selection

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

8.5. Study period

The study period will be from 01/01/2014 to 31/12/2025 or the most recent data available for each contributing data source.

It should be noted that in IQVIA LPD Belgium, the availability of the accurate data begins in February 2016, and in 2015 in FinOMOP-THL.

8.6. Variables

8.6.1. Exposure

All objectives:

Three classes of medication are of interest: HMG-CoA reductase inhibitors (statins), GLP-1 receptor agonists, and SGLT2 inhibitors.

Among these three classes, statins have been further grouped into two subclasses: hydrophilic statins and lipophilic statins. GLP-1 receptor agonists have been further grouped into three subclasses: newer generation GLP-1 injection, oral semaglutide, and earlier GLP-1. No subclass is defined for SGLT2.

The included ingredients of each drug class and subclass are listed below.

- Statins:
 - Hydrophilic statins: Rosuvastatin, pravastatin
 - Lipophilic statins: Atorvastatin, fluvastatin, lovastatin, pitavastatin, simvastatin
- GLP-1 receptor agonists:
 - Newer GLP-1 injection: Semaglutide (injection), Tirzepatide
 - Oral semaglutide
 - Earlier GLP-1: Liraglutide, Dulaglutide, Exenatide, Lixisenatide
- SGLT2 inhibitors: Dapagliflozin, Canagliflozin, Empagliflozin, Ertugliflozin

Exposure to statins, GLP-1 receptor agonists, or SGLT2 inhibitors will be identified as the first prescription or dispensing recorded, defined as no previous exposure to statins, GLP-1 receptor agonists, or SGLT2 inhibitors, respectively. Therefore, only the first-ever initiated medication in each medication class will be studied, and the subclass (for statins and GLP-1 only) will then be identified.

Treatment episodes will be generated by concatenating sequential prescriptions that have a maximum of 90 days between the end date of one prescription and the start date of the next prescription (i.e. grace window). If a sequential prescription starts before the end of the previous prescription, the overlapping days will be discarded, and the end date of the sequential prescription will depend only on the start date of the sequential prescription. Switching between medications within the same subclass (e.g. from lovastatin to simvastatin) will be allowed, while switching between subclass (e.g. from lovastatin to pravastatin) will not be allowed.

Medications will be identified using Anatomical Therapeutic Chemical (ATC) and Medical prescription normalised (RxNorm) codes. The preliminary concept sets used for the identification of exposures are described in [Annex IV](#). These codes will be refined during the study execution, following the standard DARWIN EU® phenotyping processes, which involve the review of code lists by clinical experts and the review of phenotypes after their execution in the participating data sources.

8.6.2. Outcome

Objective 2:

The following outcomes based on exposure of interests will be defined:

HMG-CoA reductase inhibitors:

- Muscle symptoms: myalgia, myopathy, rhabdomyolysis.[4,5]
- New-onset Type 2 Diabetes Mellitus (T2DM)[6]

GLP-1 receptor agonists:

- Gastro-intestinal adverse events: nausea, vomiting, delayed gastric emptying up to overt gastroparesis[7]
- Gallbladder disease (cholelithiasis, cholecystitis)
- Pancreatitis[7,8]
- Sarcopenia[9]

SGLT2 inhibitors:

- Genito-urinary events[10,11]
- Diabetic Ketoacidosis, Euglycemic Diabetic Ketoacidosis [12,13]

Muscle symptoms, gastro-intestinal adverse events, and gallbladder diseases will be assessed as combined groups, as well as individual outcomes.

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

8.6.3. Covariates

All objectives

The covariates will be as follows:

- Sex
 - Male/Female
- Age/age groups defined at the index date, namely:
 - 18 to 44
 - 45 to 54
 - 55 to 64
 - 65 to 74
 - 75+
- Calendar year of treatment initiation (assess on the index date)
- Co-medications (selected based on the indications of study medications and potential risk factors for study outcomes) as listed in **Table 2**.
 - Assessment window at the index date: [-365, -1]

Table 2. Co-medications of interest for characteristics for each of the study treatment initiator cohorts.

Medication class	ATC code(s)	Statin	GLP-1RA	SGLT2	Notes
Study medications					
Statins	C10AA		x	x	
GLP-1 receptor agonists	A10BJ	x		x	
SGLT2 inhibitors	A10BK	x	x		
Other co-medications					
Angiotensin-Converting Enzyme inhibitors (ACE inhibitors)	C09AA, C09BA, C09BB	x	x	x	Renin-angiotensin-aldosterone system (RAAS) inhibitor
Amiodarone	C01BD01	x			Statin interaction (myopathy)
Angiotensin II receptor blockers (ARBs)	C09CA, C09DA, C09DB	x	x	x	RAAS inhibitor
Anticholinergics	A03A, G04BD, N04A*		x		Delay gastric emptying
Anticoagulants	B01AA, B01AE, B01AF	x	x	x	Warfarin, Apixaban, Rivaroxaban, Edoxaban, Dabigatran
Antiplatelets	B01AC	x	x	x	Aspirin, clopidogrel, prasugrel, ticagrelor
Antipsychotics	N05A		x		Delay gastric emptying
Azole antifungals (systemic)	J02AC	x			CYP3A4 inhibitors
Beta blockers	C07	x	x	x	
Calcium Channel Blockers (CCBs)	C08	x	x	x	
Cyclosporine	L04AD01	x		x	Statin interaction; immunosuppressant

Medication class	ATC code(s)	Statin	GLP-1RA	SGLT2	Notes
Dipeptidyl peptidase-4 (DPP-4) inhibitors	A10BH		x	x	Type 2 diabetes treatment
Diuretics	C03			x	
Diltiazem	C08DB01	x			Statin interaction
Dopamine agonists	N04BC		x		Delay gastric emptying
Ezetimibe	C10AX09	x			Lipid-lowering therapy
Fibrates	C10AB	x			Increase statin-associated myopathy risk
Glucocorticoids (systemic)	H02AB	x			Myopathy risk
Immunosuppressants (other)	L04A			x	Includes biologics, antimetabolites, etc.
Insulin	A10A		x	x	Diabetes treatment
Macrolides	J01FA	x	x		CYP3A4 inhibition; affect GI motility
Meglitinides	A10BX02, A10BX03		x	x	Type 2 diabetes treatment; Repaglinide, nateglinide
Metformin	A10BA02		x	x	Type 2 diabetes treatment
Metoclopramide	A03FA01		x		Prokinetic; marker of gastroparesis/nausea
Opioids	N02A		x		Delay gastric emptying
Proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors	C10AX13, C10AX14, C10AX16	x			Blood cholesterol control: Alirocumab, evolocumab, inclisiran
Protease inhibitors	J05AE	x			Strong CYP3A4 inhibitors
Proton pump inhibitors	A02BC		x		May reflect underlying GI disease
Sulfonylureas	A10BB		x	x	Type 2 diabetes treatment
Tacrolimus	L04AD02	x		x	Statin interaction; immunosuppressant
Thiazolidinediones	A10BG		x	x	Type 2 diabetes treatment
Tricyclic antidepressants	N06AA		x		Delay gastric emptying
Verapamil	C08DA01	x			Statin interaction

ATC = Anatomical Therapeutic Chemical. GLP-1RA = Glucagon-like peptide-1 receptor agonists; SGLT2 = Sodium-glucose co-transporter 2 inhibitors; PCSK9 = Proprotein convertase subtilisin/kexin type 9; RAAS = Renin-angiotensin-aldosterone system.

x: results will be presented for the drug class.

- Comorbidities and potential indications
 - Assessment window at index date: [-Inf, 0]
 - Dyslipidaemia (composed of hypercholesterolaemia and hypertriglyceridaemia)
 - Obesity
 - Diabetes
 - Heart failure

- Stroke
- Coronary Artery Disease
- Myocardial infarction
- Chronic kidney disease (CKD)
- Hypertension

These conditions will be defined using the phenotypes developed within the DARWIN EU® Study P4-C3-006/ P4-C2-020/019 - Population demographics and disease frequency across the DARWIN EU® Network ([EUPAS1000000962](#)).

- Potential risk factors for study outcomes as listed in **Table 3**.
 - Assessment window at index date: [-Inf, 0]

Table 3. Conditions/procedures that are potential risk factors to study outcomes.

Outcome	Risk factor / Disease class	Concept ID*
Statin-induced myopathy	Chronic liver disease	Existing
	Cushing syndrome	4128029
	Osteoarthritis	80180
	Rheumatoid arthritis	Existing
	Hypothyroidism	Existing
	Hyperthyroidism	4142479
	Alcohol-related disorders	Existing
GLP-1RA-induced gastric issues	Bariatric operative procedure	4326683
	Complication due to diabetes mellitus	442793
	Nonulcer dyspepsia	4289526
	GERD	Existing
	Anal cancer and colorectal cancer	Existing
	Upper gastrointestinal tract cancer	1075457
	Peptic ulcer	4027663
GLP-1RA-induced pancreatitis	Parkinson disease	Existing
	Alcohol-related disorders	Existing
	Cholelithiasis	Study outcome
GLP-1RA-induced sarcopenia	Cholecystitis	Existing
	Chronic obstructive pulmonary disease (COPD)	Existing
	Cancer excluding non-melanoma skin cancer	Existing
SGLT2-induced urinary tract infections	Immunosuppression (HIV/AIDS, cancer, autoimmune diseases, use of antineoplastic agents, other intrinsic immune conditions)	Existing
	Disorder of lower urinary tract	4301471

*Concept ID and descendants will be used to define these conditions/procedures.

Existing: Conditions will be defined using the phenotypes developed within the DARWIN EU® Study P4-C3-006/ P4-C2-020/019 - Population demographics and disease frequency across the DARWIN EU® Network ([EUPAS1000000962](#)) or other previous studies.

GLP-1RA = Glucagon-like peptide-1 receptor agonists; SGLT2 = Sodium-glucose co-transporter 2 inhibitors; COPD = Chronic obstructive pulmonary disease; GERD = Gastroesophageal reflux disease.

The preliminary concept sets used for the identification of covariates are described in [Annex IV](#). 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.

8.7. Study size

No sample size has been calculated, as this is a descriptive disease epidemiology study and will not test a specific hypothesis. Thus, the sample size is driven by the availability of data for individuals with exposures and outcomes. Based on a preliminary feasibility assessment, the expected number of person-counts for statins in the data sources included in this study range from 150,200 in IQVIA LPD Belgium to 2,900,300 in CPRD GOLD.

The expected number of person-counts for GLP-1 in the data sources included in this study range from 10,600 in IQVIA LPD Belgium to 310,000 in FinOMOP-THL.

The expected number of person-counts for SGLT2 in the data sources included in this study range from 11,500 in IQVIA LPD Belgium to 333,200 in FinOMOP-THL.

8.8. Analysis

8.8.1. Federated network analyses

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

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

8.8.2. Data privacy protection

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

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

Objective 1

The characterisation will be conducted based on OMOP CDM mapped data using the R packages *CohortCharacteristics* and *DrugUtilisation*, developed by DARWIN EU®.[14,15]

First, initiator cohort of statins (two cohorts), GLP-1 receptor agonists (three cohorts), or SGLT2 inhibitors (one cohort) will be constructed based on the subclass of medications.

For each of the initiator cohorts, the number and proportion of individuals initiating specific drug ingredients will be reported.

The baseline characteristics, based on the variables outlined in [Section 8.6.3](#), will be described based on the following windows (days):

- Co-mediations: [-365, -1]
- Comorbidities and potential indications: [-Inf, 0]
- Potential risk factors of study outcomes: [-Inf, 0]

The co-mediations and potential risk factors of study outcomes are medication class specific and will only be reported in relevant medication classes.

Characteristics will be assessed separately for males and females, and standardised mean differences between the two groups will be reported for each characteristic.[16]

Objective 2

For the “on-treatment” follow-up, to describe the length of index treatment episodes, the distribution of: (1) number of prescriptions/dispensations concatenated under the episode; (2) sum of supply days from all prescriptions/dispensations concatenated; (3) length (in days, including grace periods) will be reported as median and inter-quartile range of each quantity.

The following adverse event incidence measures will be estimated using both “ever-treated” and “on-treatment” follow-up:

- Cumulative incidence proportions during the 5-year follow-up derived from Kaplan-Meier survival estimates with 95% confidence intervals calculated using Greenwood's formula for the variance.[16]
- Incidence rates during the 5-year follow-up, presented per 100,000 person-years, with 95% Poisson confidence intervals.

Stratification and subgroups:

All analyses will be stratified by sex and drug sub-class initiated at the index date. In a supplementary analysis incidence rate estimation will also be stratified by age groups as defined in [section 8.6.3](#).

Covariates.

The incidence rates and cumulative incidence will be calculated based on OMOP CDM mapped data using the *IncidencePrevalence* R package, developed by DARWIN EU®.[17]

8.8.4. Output

Output will be a PDF report, including an executive summary and the following tables and figures for each treatment subclass studied:

- Table 1. Baseline characteristics of treatment subclass initiators, by sex and data source.
- Figures 1–8. Sex-specific incidence rates of study outcomes among subclass initiators, by data source.
- Figures 9–16. Sex-specific cumulative incidence plots of study outcomes among subclass initiators, overall age group, by data source.
- Supplementary materials: Tables of sex- and age-group-specific incidence rates of study outcomes, by data source.

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

The MOCK tables and figures of the study output, based on simulated data, are presented in [ANNEX VI](#).

8.9. Evidence synthesis

Results from analyses described in [Section 8.8.3](#) will be presented separately for each data source. No meta-analysis of results will be conducted.

9. STRENGTHS AND LIMITATIONS

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. Misclassification of the index date of initiation may exist. Dose information is not included in this study.

Additionally, the results estimated from this study will only reflect the populations from the included data sources. Electronic health records (EHR) have certain inherent limitations because they were collected for clinical purposes rather than primarily for research use. Consequently, using 6 data sources from Belgium, Finland, Germany, the Netherlands, Spain, and the United Kingdom limits generalisability to those countries.

Study specific limitations

While possible drug combinations are included in the characterisation (e.g. dapagliflozin and metformin), these combinations will not be separated from the class-wise analysis. The metformin component in the example will be described as co-medication under objective 1.

The documentation of comorbidities, necessary for characterisation, may vary across data sources.

The on-treatment follow-up will be defined using the length of prescriptions. However, this information is not always available in the participating data sources. For example, in FinOMOP-THL, the actual end date of drug prescription and dispensing records are not available, and an assumption of 30-day exposure has been applied to both prescription and dispensing records. In IQVIA DA Germany, treatment duration is calculated as the prescribed amount divided by the Defined Daily Dose (DDD). For drug records where DDD is missing, the most frequent daily dosage for the drug is used (oral solid forms) or the assumed average maintenance dose per day (other forms).

For GLP-1, routes (oral or injection) will be defined through the recorded route in each data source. However, this information may not always be available, and misclassification of routes could exist. In this study, oral semaglutide will be identified as those specifically recorded as oral route. Records that either state injection or non-specific route will be included in the injection route. If individual has records of both oral route and injection route on the same date, the individual will be excluded.

Data source specific limitations

In IQVIA LPD Belgium and IQVIA DA Germany, the observation period of the patients is inferred based on the last visit, observation, or interaction of the individual with the health care system. This methodology impacts identifying individuals “at risk” of experiencing the outcome (i.e. the individuals included in the denominator populations) during the latest months of available data. Healthy and/or non-frequent users of the health care system will no longer be considered under observation after their last recorded visit, which might be far from the end of data, thus not contributing to denominators toward the end of data availability. Consequently, incidence rate denominators will be artificially smaller, leading to the overestimation of incidence rates. To reduce the effect of the artifacts in the denominator, the last 3 months of data in IQVIA LPD Belgium and IQVIA DA Germany will be removed for all the study objectives.

10. REFERENCES

- 1 Mosca L, Benjamin EJ, Berra K. Effectiveness-Based Guidelines for the Prevention of Cardiovascular Disease in Women-2011 Update. *Circulation*. 2011;123:1243–62.

- 2 Regitz-Zagrosek V, Oertelt-Prigione S, Prescott E. Gender differences in the effect of cardiovascular drugs: a position document of the Working Group on Pharmacology and Drug Therapy of the European Society of Cardiology. *European Heart Journal*. 2016;37:267–307.
- 3 Tamargo J, Delpón E. Sex-related differences in cardiovascular pharmacotherapy: fiction or fact? Why can't we see the evidence? *Eur Heart J Cardiovasc Pharmacother*. 2025;11:638–52.
- 4 Karalis DG, Wild RA, Maki KC, *et al*. Gender differences in side effects and attitudes regarding statin use in the Understanding Statin Use in America and Gaps in Patient Education (USAGE) study. *J Clin Lipidol*. 2016;10:833–41.
- 5 Cangemi R, Romiti GF, Campolongo G, *et al*. Gender related differences in treatment and response to statins in primary and secondary cardiovascular prevention: The never-ending debate. *Pharmacol Res*. 2017;117:148–55.
- 6 Kao DP, Martin JL, Aquilante CL, *et al*. Sex-differences in reporting of statin-associated diabetes mellitus to the US Food and Drug Administration. *BMJ Open Diabetes Res Care*. 2024;12:e004343.
- 7 Sodhi M, Rezaeianzadeh R, Kezouh A, *et al*. Risk of gastrointestinal adverse events associated with glucagon-like peptide-1 receptor agonists for weight loss. *JAMA*. 2023;330:1795–7.
- 8 Chiang C-H, Jaroenlapnopparat A, Colak SC, *et al*. Glucagon-like peptide-1 receptor agonists and gastrointestinal adverse events: A systematic review and meta-analysis. *Gastroenterology*. 2025;169:1268–81.
- 9 Scheen AJ. GLP-1-derived therapies and sarcopenia: plea for a specific focus on at risk special populations. *Diabetes Metab*. 2026;52:101708.
- 10 Gorgojo-Martínez JJ, Górriz JL, Cebrián-Cuenca A, *et al*. Clinical recommendations for managing genitourinary adverse effects in patients treated with SGLT-2 inhibitors: A multidisciplinary expert consensus. *J Clin Med*. 2024;13:6509.
- 11 Li CX, Liu TT, Zhang Q, *et al*. Safety of sodium-glucose transporter 2 (SGLT-2) inhibitors in patients with type 2 diabetes: a meta-analysis of cohort studies. *Front Pharmacol*. 2023;14:1275060.
- 12 He Z, Lam K, Zhao W, *et al*. SGLT-2 inhibitors and euglycemic diabetic ketoacidosis/diabetic ketoacidosis in FAERS: a pharmacovigilance assessment. *Acta Diabetol*. 2023;60:401–11.
- 13 Peters AL, Buschur EO, Buse JB, *et al*. Euglycemic diabetic ketoacidosis: A potential complication of treatment with sodium-glucose cotransporter 2 inhibition. *Diabetes Care*. 2015;38:1687–93.
- 14 DrugUtilisation. <https://darwin-eu.github.io/DrugUtilisation/> (accessed 10 July 2026)
- 15 Du M, Prats-Urbe A, Mercadé-Besora N, *et al*. CohortCharacteristics: an R package for population characterisation in observational studies using the OMOP common data model. *Eur J Epidemiol*. 2026;41:507–17.
- 16 Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009;28:3083–107.
- 17 Raventós B, Català M, Du M, *et al*. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. *Pharmacoepidemiol Drug Saf*. 2024;33:e5717.

11. ANNEXES

ANNEX I. Description of data sources

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

#	Section	Description
1	Data source identification and country	IQVIA LPD Belgium (IQVIA Longitudinal Patient Database Belgium) Belgium
2	Data partner information section	IQVIA (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 – General Practitioner. 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 Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. 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 (data source specific)	No QC. Integrity constraints only.
10	Linkage	No linkage.
11	Vital status	Death information is derived from healthcare events.
12	Limitations	Observation period is defined from first to last visit, causing a drop in 'healthy' persons at the end of the interval. This creates an artefactual increase in incidence and prevalence ratios. There is no sociodemographic or over-the-counter medication information.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111116 Website: https://iqvia.com

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

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

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

Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)

#	Section	Description
1	Data source identification and country	IQVIA DA Germany (IQVIA Disease Analyzer Germany) Germany
2	Data partner information section	IQVIA (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 – General Practitioner, and primary care specialists (e.g. paediatricians). Diagnoses, medication, and procedures from an ambulatory setting. Medications are recorded as prescriptions of marketed products.

#	Section	Description
5	Data collection process	Outpatient electronic health records. By clinicians at healthcare contact.
6	General representativeness	No specific details on general representativeness given.
7	Data content / source coding	Prescription is on product code level (German PZN), ICD10, NFC, Local lab coding.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. There can be patients registered under different ID numbers, because there is no linkage between different GPs.
9	Quality control (data source specific)	Data is quality checked on plausibility.
10	Linkage	No.
11	Vital status	Death information is derived from medical events.
12	Limitations	Observation period is defined from first to last visit, causing a drop in 'healthy' persons at the end of the interval. This creates an artefactual increase in incidence and prevalence ratios.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104282 Website: https://www.iqvia.com/

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

#	Section	Description
1	Data source identification and country	SIDIAP (The Information System for Research in Primary Care) Catalunya, Spain
2	Data partner information section	Institut d'Investigació en Atenció Primària de Salut Jordi Gol (IDIAPJGol)
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. SIDIAP is a database of primary care electronic health records of the population of Catalonia, North-East Spain. It contains pseudo-anonymised records of more than 8 million people, of which 6.1 million are active as of 2022, representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – General Practitioner, and hospital inpatient care. SIDIAP captured data includes routine visits, socio-demographics, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, sick leaves and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Català de la Salut (Catalan Health Institute) is the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content / source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.

#	Section	Description
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient has a unique id, also upon changing practice or re-registration.
9	Quality control (data source specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g., there are several variables with information concerning the women's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinally and reliability), plausibility (range checks and unusual values), and consistency are inspected through visualisation tools. In addition, before accessing the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences, and prevalence of the data requested for the project, allowing detection of inconsistencies in the data extraction prior to data delivery. External validation processes of the SIDIAP database mainly include assessing the data recorded in SIDIAP through linkage to external gold standard data sources, by analysing free text, or by sending questionnaires to health professionals.
10	Linkage	SIDIAP is linked to a hospital discharge database, pharmacy dispensation, and primary care laboratories. It can also be linked to other registries in Catalonia on a project by project basis.
11	Vital status	Mortality is fully captured in SIDIAP. The cause of death is not available, but can be linked to the Spanish death registry on a project by project basis.
12	Limitations	The SIDIAP data is not representative of individuals not using public primary care, and conditions that are usually followed by specialist care might not be properly captured. In addition, there is limited information on lifestyle variables. Patients are followed until Death or when transferring to another primary health care centre that does not contribute to SIDIAP.
13	Main references	Recalde M, Rodríguez C, Burn E, Far M, García D, Carrere-Molina J, Benítez M, Moleras A, Pistillo A, Bolívar B, Aragón M, Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

The Netherlands: Integrated Primary Care Information (IPCI)

#	Section	Description
1	Data source identification and country	IPCI (Integrated Primary Care Information) the Netherlands
2	Data partner information section	Erasmus University Medical Center (EMC) Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards.

#	Section	Description
		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 – General Practitioner. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. The patientdossiers are collected by Erasmus MC data managers and combined in one harmonized database. Several checks are done on this database to ensure correct data processing. Persons can have dossiers at multiple GPs.
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content / source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients can be registered under different IDs, but since a patient can only be registered at one GP at a time, the observations periods will not overlap.
9	Quality control (data source specific)	Prior to each data release, extensive quality control steps are performed, e.g., comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g. number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical data, e.g. distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Capture of oral contraceptives in IPCI is incomplete. The first prescription is recorded within the patient's medical file but not the subsequent prescriptions. If you want to use IPCI data for research questions on oral contraceptives, you can do that but then you need to select those GPs that use MEDICOM as software system (where you have prescribing and dispensing).

#	Section	Description
		However, of hormone substitution therapy, the information on prescribing is COMPLETE. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.
13	Main references	de Ridder MAJ, de Wilde M, de Ben C, Leyba AR, Mosseveld BMT, Verhamme KMC, van der Lei J, Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182144
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

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

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

#	Section	Description
		However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (data source specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (UTS). Each practice is associated to an UTS date set when the data quality standards become satisfactory, and CPRD recommend using only longitudinal data starting from this UTS date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards, and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice UTS date. CPRD also checks data quality standards at the patient level, and associates each patient with a flag, reporting if its data are acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at the Practice Level.
11	Vital status	The date of death in CPRD GOLD has been validated against the Population registry (ONS) mortality data. Reference: https://doi.org/10.1002/pds.4747
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL, Dedman D, Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://www.cprd.com/data/primary-care-data/cprd-gold

ANNEX II. Fitness for use assessment

Data source justification for inclusion and key characteristics

Out of 49 Data Partners currently onboarded into the DARWIN EU® Network of Data Partners (as of 02/06/2026), 28 cover the required setting for the study (primary care). 17 of these cover the time period 2017–2024. Of these, Clinical Practice Research Datalink Aurum (CPRD Aurum), Croatian National Public Health Information System (NAJS), Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT), and Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (Pharmacoepidemiological Research Database for Public Health Systems, BIFAP) were deprioritised, as there would be issues with turnaround timing. Valencia Health System Integrated Dataset (VID) was also excluded, as it does not include females born before 1953. Papageorgiou General Hospital (PGH) and Hospital Universitario Virgen Macarena (HUVIM) are hospital-based data sources, and although they declare to report general practitioner data, this is only true for hospitalised patients; therefore, PGH and HUVIM were also excluded. Lithuanian Digital Health Platform (LDH) was only recently onboarded, and therefore would require more time for study execution, which could compromise the short timelines of this study. Ultimately, CPRD GOLD, FinOMOP-THL, Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED), IPCI, IQVIA DA Germany, IQVIA LPD Belgium, InGef Research Database (InGef RDB), and SIDIAP were selected for this study.

Both InGef RDB and IQVIA DA Germany have low counts of muscle symptoms. Only 2 data sources have >100 counts for Sarcopenia (HI-SPEED, SIDIAP) but the remainder have enough counts for the other outcomes. IPCI and IQVIA LPD Belgium have no counts of Ketoacidosis. HI-SPEED was not able to participate in the study due to limited capacity. Therefore, the following 6 data sources were ultimately selected:

1. IQVIA LPD Belgium
2. FinOMOP-THL
3. IQVIA DA Germany
4. SIDIAP
5. CPRD GOLD
6. IPCI

IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)

IQVIA LPD Belgium will be included in this study because it is a primary care data source that provides relevant information on the medications and outcomes of interest in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for statins, GLP-1 receptor agonists, and SGLT2 inhibitors (all objectives) in IQVIA LPD Belgium will be 150,200; 10,600; and 11,500, respectively.

Moreover, data availability and follow-up in IQVIA LPD Belgium is sufficient, as data availability starts in 2016, and the date of the most recent data extraction is 03/2026, which aligns with the study period.

There are some study specific limitations present in IQVIA LPD Belgium, namely: the observation period of the patients is calculated based on the last visit, observation, or interaction of the individual with the health care system.

This affects which individuals are considered “at risk” for the outcomes of interest (i.e. included in the denominator), particularly in the most recent months of data, where healthier or infrequent users may no longer appear active. As a result, the denominator used to calculate the incidence of outcomes of interest may show an artificial decrease while the numerator (incident user) remains. Therefore, the last three

months of data will be excluded from the analysis to mitigate the impact. Due to data availability, the study period for IQVIA LPD Belgium will be 2016–2025.

Lastly, IQVIA LPD Belgium has blanket Institutional Review Board (IRB) approval, which makes the execution of this study feasible within the current study timelines.

Finnish Care Register for Health Care (FinOMOP-THL)

FinOMOP-THL will be included in this study because it is a primary care data source that provides relevant information on the medications and outcomes of interest in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for statins, GLP-1 receptor agonists, and SGLT2 inhibitors (all objectives) in FinOMOP-THL will be 2,238,900; 310,000; and 333,200, respectively.

Moreover, data availability and follow-up in FinOMOP-THL is sufficient, as data availability starts in 2015, and the date of the most recent data extraction is 12/2025, which aligns with the study period.

There are some study specific limitations present in FinOMOP-THL, namely: Due to data availability, the study period for FinOMOP-THL will be 2015–2025.

Actual end date of drug prescription and dispensing records are not available in FinOMOP-THL, and an assumption of 30-day exposure has been applied to both prescription and dispensing records, which could lead to treatment episodes ending earlier.

Lastly, IRB approval for FinOMOP-THL is estimated to take 3 months, which makes the execution of this study feasible within the current study timelines.

IQVIA Disease Analyzer Germany (IQVIA DA Germany)

IQVIA DA Germany will be included in this study because it is a primary care data source that provides relevant information on the medications and outcomes of interest in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for statins, GLP-1 receptor agonists, and SGLT2 inhibitors (all objectives) in IQVIA DA Germany will be 1,397,300; 118,300; and 271,300, respectively.

Moreover, data availability and follow-up in IQVIA DA Germany is sufficient, as data availability starts in 1993, and the date of the most recent data extraction is 06/2025, which aligns with the study period.

There are some study specific limitations present in IQVIA DA Germany, namely: the observation period of the patients is calculated based on the last visit, observation, or interaction of the patient with the health care system. Due to data availability, the study period for IQVIA DA Germany will be 2014–03/2026. The last three months of data will be excluded.

Treatment duration in IQVIA DA Germany is calculated as the prescribed amount divided by the DDD. For drug records where DDD is missing, the most frequent daily dosage for the drug is used (oral solid forms) or the assumed average maintenance dose per day (other forms).

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

Information System for Research in Primary Care (SIDIAP)

SIDIAP will be included in this study because it is a primary care data source that provides relevant information on the medications and outcomes of interest in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for statins, GLP-1 receptor agonists, and SGLT2 inhibitors (all objectives) in SIDIAP will be 1,945,800; 80,100; and 259,300, respectively.

Moreover, data availability and follow-up in SIDIAP is sufficient, as data availability starts in 2006, and the date of the most recent data extraction is 12/2024, which aligns with the study period.

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

Integrated Primary Care Information (IPCI)

IPCI will be included in this study because it is a primary care data source that provides relevant information on the medications and outcomes of interest in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for statins, GLP-1 receptor agonists, and SGLT2 inhibitors (all objectives) in IPCI will be 453,100; 28,400; and 33,100, respectively.

Moreover, data availability and follow-up in IPCI is sufficient, as data availability starts in 2006, and the date of the most recent data extraction is 03/2026, which aligns with the study period.

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

Clinical Practice Research Datalink GOLD (CPRD GOLD)

CPRD GOLD will be included in this study because it is a primary care data source that provides relevant information on the medications and outcomes of interest in the general population.

Based on a preliminary feasibility assessment, the expected number of person-counts for statins, GLP-1 receptor agonists, and SGLT2 inhibitors (all objectives) in CPRD GOLD will be 2,900,300; 75,500; and 140,900, respectively.

Moreover, data availability and follow-up in CPRD GOLD is sufficient, as data availability starts in 1989, and the date of the most recent data extraction is 11/2025, which aligns with the study period.

Lastly, CPRD GOLD has blanket IRB approval in principle, but an amendment is required for this study. The amendment process is estimated to take 2 months, which makes the execution of this study feasible within the current study timelines.

ANNEX III. Operational and reporting considerations

DATA MANAGEMENT

Data management

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

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

Data storage and protection

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

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

The output files are stored in the DARWIN EU® Digital Research Environment (DRE). These output files do not contain any data that allow identification of subjects included in the study. The DRE implements further security measures to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/2016 in the various member states.

QUALITY CONTROL

Data source quality control

When defining drug cohorts, non-systemic products will be excluded from the list of included codes summarised on the ingredient level.

When defining cohorts for indications, a systematic search of possible codes for inclusion will be identified using the *CodelistGenerator* R package (<https://github.com/darwin-eu/CodelistGenerator>). This package allows the user to define a search strategy and will use this to query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, the *CohortDiagnostics* (<https://github.com/OHDSI/CohortDiagnostics>) and *DrugExposureDiagnostics* (<https://cran.r-project.org/web/packages/DrugExposureDiagnostics/index.html>) R packages will be run, if needed, to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error. The *DrugExposureDiagnostics* package evaluates ingredient-specific attributes and patterns in drug exposure records.

The study code will be based on DARWIN EU® R packages: *IncidencePrevalence* to estimate Incidence and Prevalence, and *CohortCharacteristics* to characterise the cohort by indication. These packages will include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub.

PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

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

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

ANNEX IV. Preliminary code list for exposures and outcomes

Table S1. Preliminary list of conditions definitions.

Concept name	Concept ID (including descendants)	Vocabulary
Myositis	73001	SNOMED
Muscle pain	442752	SNOMED
Rhabdomyolysis	4345578	SNOMED
Diabetes mellitus	201820	SNOMED
Nausea	31967	SNOMED
Gastroparesis syndrome	195847	SNOMED
Vomiting	441408	SNOMED
Delayed gastric emptying	4198430	SNOMED
Cholecystitis	192956	SNOMED
Biliary calculus	4145627	SNOMED
Pancreatitis	4192640	SNOMED
Sarcopenia	36674346	SNOMED
Urinary tract infectious disease	81902	SNOMED
Female genital tract infection	4129391	SNOMED
Diabetic ketoacidosis	443727	SNOMED

Table S2. Preliminary list of medicines definitions.

Medication class	Concept name	ATC code	Ingredient Concept ID
Statins	rosuvastatin	C10AA07	1510813
Statins	simvastatin	C10AA01	1539403
Statins	atorvastatin	C10AA05	1545958
Statins	fluvastatin	C10AA04	1549686
Statins	pravastatin	C10AA03	1551860
Statins	lovastatin	C10AA02	1592085
Statins	pitavastatin	C10AA08	40165636
GLP-1	semaglutide	A10BJ06	793143
GLP-1	exenatide	A10BJ01	1583722
GLP-1	liraglutide	A10BJ02	40170911
GLP-1	lixisenatide	A10BJ03	44506754
GLP-1	dulaglutide	A10BJ05	45774435
GLP-1	Tirzepatide	A10BX16	779705
SGLT2	ertugliflozin	A10BK04	793293
SGLT2	canagliflozin	A10BK02	43526465
SGLT2	dapagliflozin	A10BK01	44785829
SGLT2	empagliflozin	A10BK03	45774751

ATC= Anatomical Therapeutic Chemical.

ANNEX V. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Sex-specific incidence of selected adverse events for major cardio-metabolic drug classes

EU PAS Register® number: EUPAS1000001081

Study reference number (if applicable): P5-C1-008/009/010

Section 1: Milestones	Yes	No	N/A	Section Number
1.1 Does the protocol specify timelines for				8.1 – 8.3
1.1.1 Start of data collection ¹	☒	☐	☐	
1.1.2 End of data collection ²	☒	☐	☐	
1.1.3 Progress report(s)	☐	☐	☒	
1.1.4 Interim report(s)	☐	☐	☒	
1.1.5 Registration in the EU PAS Register®	☐	☐	☒	
1.1.6 Final report of study results.	☒	☐	☐	

Comments:

Section 2: Research question	Yes	No	N/A	Section Number
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	6
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)	☒	☐	☐	6
2.1.2 The objective(s) of the study?	☒	☐	☐	7
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	8.3
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no <i>a priori</i> hypothesis?	☐	☐	☒	

Comments:

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

² Date from which the analytical dataset is completely available.

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

Comments:

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

Comments:

Section 5: Exposure definition and measurement		Yes	No	N/A	Section Number
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	8.6.1
5.2	Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☐	☒	
5.3	Is exposure categorised according to time windows?	☒	☐	☐	8.8

<u>Section 5: Exposure definition and measurement</u>	Yes	No	N/A	Section Number
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☒	☐	☐	8.8
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	
5.6 Is (are) (an) appropriate comparator(s) identified?	☐	☐	☒	

Comments:

<u>Section 6: Outcome definition and measurement</u>	Yes	No	N/A	Section Number
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	8.8.3
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.8.3
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☐	☒	
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

<u>Section 7: Bias</u>	Yes	No	N/A	Section Number
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☐	☐	☒	
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☐	☐	☒	
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	9

Comments:

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

Comments:

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

Comments:

<u>Section 10: Analysis plan</u>	Yes	No	N/A	Section Number
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.8.3
10.2 Is study size and/or statistical precision estimated?	☐	☐	☒	
10.3 Are descriptive analyses included?	☒	☐	☐	8.8.3

<u>Section 10: Analysis plan</u>	Yes	No	N/A	Section Number
10.4 Are stratified analyses included?	☒	☐	☐	8.8.3
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?	☒	☐	☐	8.3
10.8 Are relevant sensitivity analyses described?	☒	☐	☐	8.8.3

Comments:

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

Comments:

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

Comments:

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

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

Comments:

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

Comments:

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

Comments:

Name of the main author of the protocol: Xintong Li

Date: 10/07/2026

Signature: Xintong Li

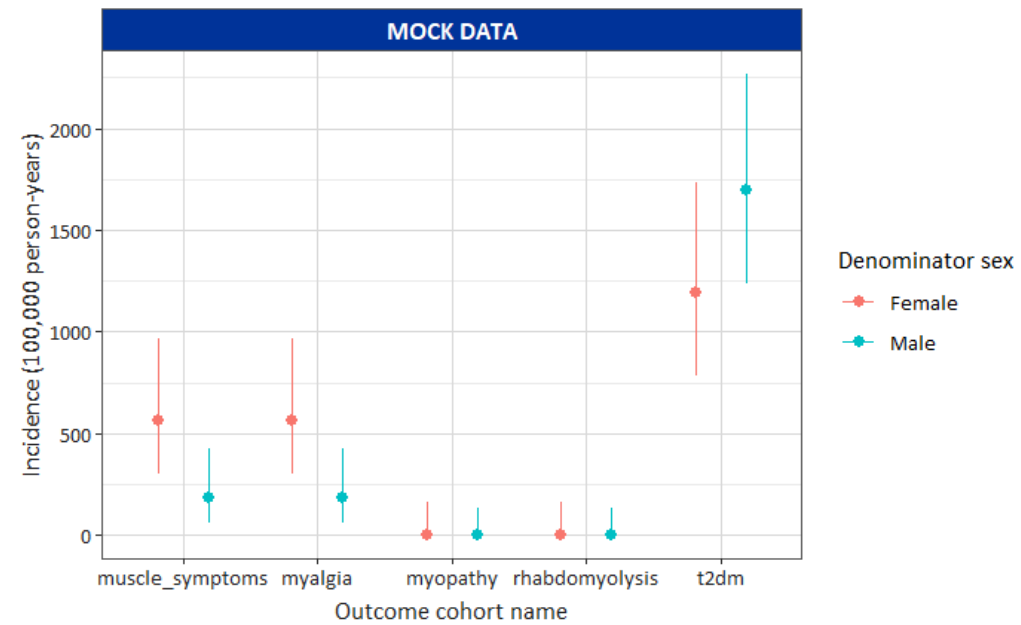
ANNEX VI. Mock tables and figures

The following MOCK tables and figures use statins initiators for demonstration. Same tables and figures will be generated for the study report for GLP-1 and SGLT2 results.

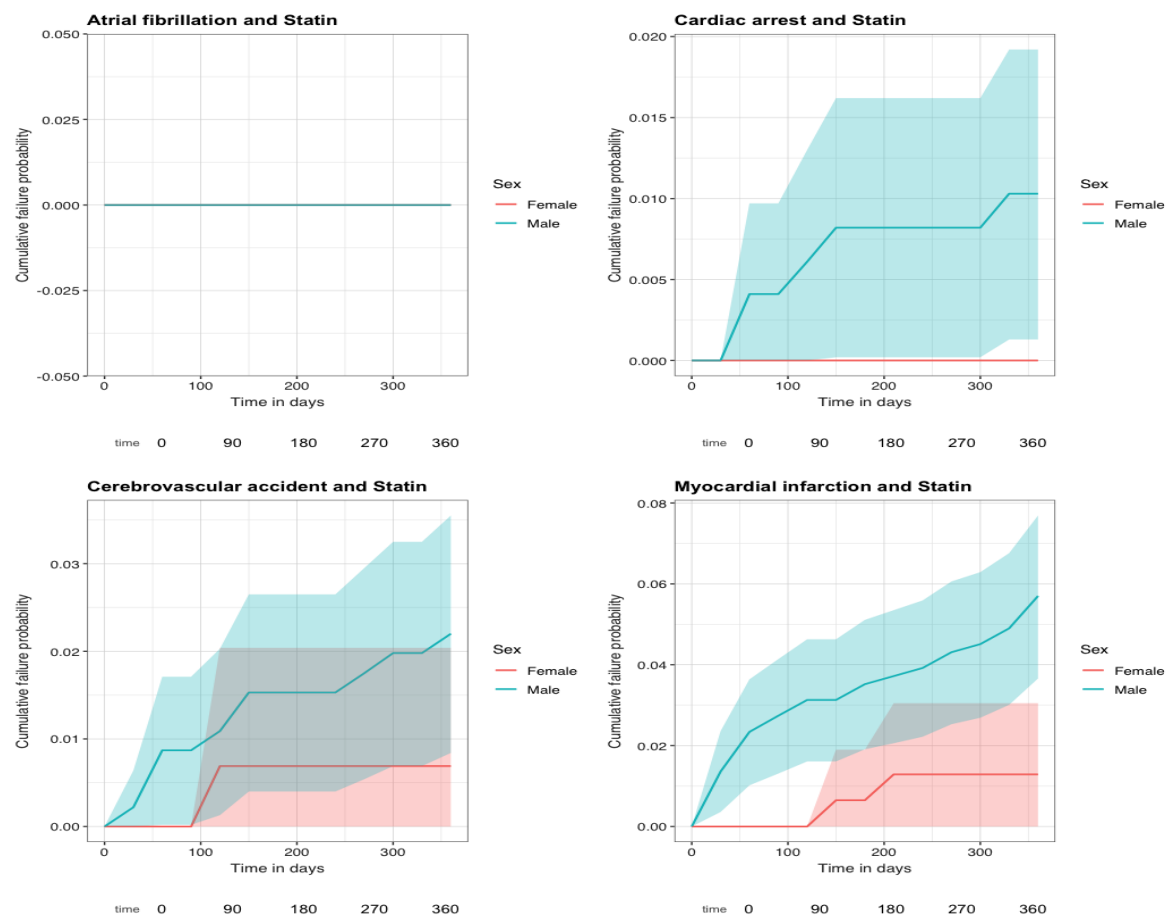
MOCK Table 1. Baseline characteristics of lipophilic statins initiators, by data source.

Variable name	Variable level	Estimate name	Data Source									
			Data Source A		Data Source B		Data Source C		Data Source D		Data Source E	
			Sex									
			Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Number subjects	–	N	179	640	179	640	179	640	179	640	179	640
Age group	18 to 44	N (%)	10 (5.59%)	40 (6.25%)	10 (5.59%)	40 (6.25%)	10 (5.59%)	40 (6.25%)	10 (5.59%)	40 (6.25%)	10 (5.59%)	40 (6.25%)
	45 to 54	N (%)	20 (11.17%)	80 (12.50%)	20 (11.17%)	80 (12.50%)	20 (11.17%)	80 (12.50%)	20 (11.17%)	80 (12.50%)	20 (11.17%)	80 (12.50%)
	55 to 64	N (%)	30 (16.76%)	120 (18.75%)	30 (16.76%)	120 (18.75%)	30 (16.76%)	120 (18.75%)	30 (16.76%)	120 (18.75%)	30 (16.76%)	120 (18.75%)
	65 to 74	N (%)	40 (22.35%)	160 (25.00%)	40 (22.35%)	160 (25.00%)	40 (22.35%)	160 (25.00%)	40 (22.35%)	160 (25.00%)	40 (22.35%)	160 (25.00%)
	75+	N (%)	79 (44.13%)	240 (37.50%)	79 (44.13%)	240 (37.50%)	79 (44.13%)	240 (37.50%)	79 (44.13%)	240 (37.50%)	79 (44.13%)	240 (37.50%)
Age	–	Median [Q25 – Q75]	71 [59 – 82]	68 [59 – 79]	71 [59 – 82]	68 [59 – 79]	71 [59 – 82]	68 [59 – 79]	71 [59 – 82]	68 [59 – 79]	71 [59 – 82]	68 [59 – 79]
		Mean (SD)	67.30 (19.56)	67.33 (15.45)	67.30 (19.56)	67.33 (15.45)	67.30 (19.56)	67.33 (15.45)	67.30 (19.56)	67.33 (15.45)	67.30 (19.56)	67.33 (15.45)
		Range	18 to 96	19 to 98	18 to 96	19 to 98	18 to 96	19 to 98	18 to 96	19 to 98	18 to 96	19 to 98
Statin ingredient at initiation	Atorvastatin	N (%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)

Variable name	Variable level	Estimate name	Data Source									
			Data Source A		Data Source B		Data Source C		Data Source D		Data Source E	
			Sex									
			Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
	Simvastatin	N (%)	170 (94.97%)	541 (84.53%)	170 (94.97%)	541 (84.53%)	170 (94.97%)	541 (84.53%)	170 (94.97%)	541 (84.53%)	170 (94.97%)	541 (84.53%)
	Cerivastatin	N (%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	Fluvastatin	N (%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	Lovastatin	N (%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	Pitavastatin	N (%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	Pravastatin	N (%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	Rosuvastatin	N (%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Comorbidities	Hypertension	N (%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)
	Type 2 diabetes	N (%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)
Co-medications	GLP-1		9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)	9 (5.03%)	99 (15.47%)



MOCK Figures 1–8. Sex-specific incidence rates of study outcomes among lipophilic statins initiators, overall age group, by data source.



MOCK Figures 9–16. Sex-specific cumulative incidence of study outcomes among lipophilic statins initiators.

MOCK Supplementary Table. Incidence rates per 100,000 person-years and 95% CI of study outcomes among on-treatment lipophilic statins initiator cohort.

	Data Source									
	Mock A		Mock B		Mock C		Mock D		Mock E	
	Sex									
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Myalgia; Overall	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)
Myalgia; 18–44	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)
Myalgia; 45–54	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)	7,243.69 (3,312.28 – 13,750.79)	26,277.37 (21,513.75 – 31,781.72)
Myalgia; 55–64	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)
Myalgia; 65–74	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)
Myalgia; 75+	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)	6,650.48 (805.40 – 24,023.83)	18,148.00 (10,373.15 – 29,471.21)
Myopathy; Overall	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 – 4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 – 4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 – 4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 – 4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 – 4,098.21)
Myopathy; 18–44	0.00 (0.00 – 11,486.47)	0.00 (0.00 – 4,267.31)	0.00 (0.00 – 11,486.47)	0.00 (0.00 – 4,267.31)	0.00 (0.00 – 11,486.47)	0.00 (0.00 – 4,267.31)	0.00 (0.00 – 11,486.47)	0.00 (0.00 – 4,267.31)	0.00 (0.00 – 11,486.47)	0.00 (0.00 – 4,267.31)

	Data Source									
	Mock A		Mock B		Mock C		Mock D		Mock E	
	Sex									
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Myopathy; 45–54	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)
Myopathy; 55–64	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)
Myopathy; 65–74	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)
Myopathy; 75+	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)
Type 2 diabetes; Overall	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)	1,845.53 (223.50 – 6,666.69)	1,989.06 (799.70 –4,098.21)
Type 2 diabetes; 18–44	0.00 (0.00 – 12,250.94)	0.00 (0.00 – 4,474.79)	0.00 (0.00 – 12,250.94)	0.00 (0.00 – 4,474.79)	0.00 (0.00 – 12,250.94)	0.00 (0.00 – 4,474.79)	0.00 (0.00 – 12,250.94)	0.00 (0.00 – 4,474.79)	0.00 (0.00 – 12,250.94)	0.00 (0.00 – 4,474.79)
Type 2 diabetes; 45–54	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)
Type 2 diabetes; 55–64	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)

	Data Source									
	Mock A		Mock B		Mock C		Mock D		Mock E	
	Sex									
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Type 2 diabetes; 65–74	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 – 3,265.25)
Type 2 diabetes; 75+	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 –3,265.25)	0.00 (0.00 – 3,312.93)	1,500.18 (550.54 – 3,265.25)