1.Title Page

Title	TARGET-EU: Comparative effectiveness and safety studies using the target trial emulation and estimand frameworks: Dapagliflozin and Major Adverse Cardiovascular Events in Type 2 Diabetes
Research question & Objectives	The primary objective is to assess whether dapagliflozin reduces the risk of major adverse cardiovascular events compared to DPP-4 inhibitors in patients with type 2 diabetes and established cardiovascular disease or at least two cardiovascular risk factors
Protocol version	V1.0
Last update date	27 February 2026
Contributors	**Primary investigator contact information:** Dr. Daniala Weir (d.l.weir@uu.nl) **Contributor names:** Prof. Olaf Klungel Prof. Ian Douglas Dr. Maria del Mar Martin Perez Dr. Belen Castillo Cano Dr. Elina Martin Marino Dr. Patrick Souverein
Study registration	**Site:** https://catalogues.ema.europa.eu/node/4440/administrative-details **Identifier:** EUPAS1000000539
Sponsor	**Organization:** EU PE&PV research network **Contact:** eupepv@uu.nl
Conflict of interest	None

Table of Contents

2. Abstract

Background

Dapagliflozin, a sodium-glucose co-transporter 2 (SGLT2) inhibitor, has demonstrated cardiovascular benefits in randomized trials, including reduced hospitalization for heart failure and renal disease progression. Real-world studies have shown that SGLT2 inhibitors consistently reduce hospitalization for heart failure, with some evidence of mortality benefits, but their effects on major atherosclerotic cardiovascular events (MI and stroke) are less consistent.

Objectives

The primary objective is to estimate the effect of initiating dapagliflozin versus a DPP-4 inhibitor on time to first MACE, defined as a composite of non-fatal myocardial infarction, non-fatal stroke, and all-cause mortality among patients who have or are at risk of atherosclerotic cardiovascular disease (ASCVD).

Methods

We will conduct an active comparator new-user cohort study using linked electronic health records from the UK (CPRD GOLD and Aurum) and Spain (BIFAP). Eligible individuals are adults (≥40 years) with T2D and either established ASCVD or high risk of ASCVD, who initiated dapagliflozin or a DPP-4 inhibitor between 2012 and 2024. In the primary analysis, a treatment policy strategy is used for treatment-related intercurrent events (discontinuation, switching, or intensification), and a composite strategy is applied to all-cause death. Inverse probability of treatment weighting (IPTW) is used to estimate treatment effects in the study population. The primary analysis uses a Cox proportional hazards model, with supplemental analyses using an accelerated failure time model to estimate restricted mean survival time at 3 and 5 years. Sensitivity analyses will be conducted to assess the impact of varying assumptions about the censoring-at-random condition using inverse probability of censoring weights, how the estimated treatment effect changes when the censoring mechanism deviates from censoring at random using tipping point analysis and exposure misclassification using probabilistic bias analysis.

3. Amendments and updates

Version date	Version number	Section of protocol	Amendment or update	Reason
27 February 2026	1.0			

4. Milestones

Table 1. Milestones

Milestone	Date
Study protocol for RWD study	8 August 2025
Preliminary results RWD study	April 2026
Final Study report	10 June 2026

5. Background

The purpose of this study is to evaluate the effectiveness of dapagliflozin compared to DPP-4 inhibitors, when added to usual care, in reducing major adverse cardiovascular events (MACE) among patients with type 2 diabetes who have or are at risk for atherosclerotic cardiovascular disease (ASCVD). While initial cardiovascular outcome trials (CVOTs) were primarily designed to assess the safety of glucose-lowering therapies, accumulating evidence — including from the DECLARE-TIMI 58 trial — has demonstrated that dapagliflozin not only meets cardiovascular safety standards but also confers benefits in reducing hospitalization for heart failure and slowing renal disease progression.[1] In contrast, DPP-4 inhibitors have shown a neutral cardiovascular profile, as observed in placebo-controlled CVOTs such as TECOS (sitagliptin)[2] and SAVOR-TIMI 53 (saxagliptin).[3] Importantly, no large head-to-head randomized controlled trial has directly compared dapagliflozin to a DPP-4 inhibitor for MACE or other cardiovascular outcomes, leaving a significant evidence gap regarding their comparative effectiveness. Non-interventional studies using real-world data sources have been conducted to fill this gap.

Patorno et al. demonstrated that SGLT2 inhibitors were associated with a lower risk of hospitalization for heart failure compared with either non-gliflozin therapies or GLP-1 receptor agonists across categories of cardiovascular disease, while showing no consistent difference in myocardial infarction or stroke. In Europe, registry- and primary care–based studies conducted in adults with type 2 diabetes from Scandinavian national registries and UK primary care (CPRD) have also reported reduced risks of heart failure hospitalization and mortality with SGLT2 inhibitors compared with other glucose-lowering drugs. Gonzalez et al. (2023) conducted a large Medicare-based cohort study of older adults with type 2 diabetes and heart failure, comparing SGLT2 inhibitors with DPP-4 inhibitors and GLP-1 receptor agonists. They found that initiation of SGLT2 inhibitors was associated with substantially lower risks of hospitalization for heart failure and, among patients with reduced ejection fraction, a modest but significant reduction in myocardial infarction or stroke compared with DPP-4 inhibitors. This real-world evidence complements trial findings by confirming that SGLT2 inhibitors consistently reduce heart failure hospitalizations and, in some populations, may modestly reduce ischemic events or

mortality. However, their benefits for major atherosclerotic cardiovascular events (MI, stroke) appear less consistent and may depend on patient subgroups (e.g., those with HFrEF).

This non-interventional study is therefore designed to contribute to this evidence base by providing causal, comparative evidence under real-world conditions among patients with ASCVD or those at high risk. In doing so, it aligns with the European Medicines Agency's regulatory emphasis on comparative effectiveness[4] and supports evolving clinical guidelines that prioritize SGLT2 inhibitors, like dapagliflozin, in patients with high cardiovascular risk.[5,6]

6. Research questions and objectives

The overall aim is to assess whether dapagliflozin reduces the time to first major adverse cardiovascular event compared to DPP-4 inhibitors in patients with type 2 diabetes and established cardiovascular disease or at least two cardiovascular risk factors.

6.1 Primary Estimand 1

Research question targeted by the estimand 1: What is the HR of MACE for Dapagliflozin vs DPP-4i in patients with type 2 diabetes with or at risk for ASCVD regardless of treatment discontinuation, switching or new add-on antihyperglycemic therapy?

Table 2. Estimand 1

Attribute	Target Trial	Target Trial Emulation	Comment
Population	Patients with type 2 diabetes who have or are at risk for ASCVD **Established ASCVD:** history or diagnosis of ischemic heart disease, ischemic cerebrovascular disease, or peripheral arterial disease before the index date. **High ASCVD risk**: no established ASCVD, but age ≥55 (men) or ≥60 (women) plus at least one of the following: dyslipidemia, lipid-lowering therapy, hypertension, antihypertensive therapy, or tobacco use within the past year.	Same, but population identified using RWD (primary care, hospital records, prescription records).	Potential for mismeasurement of tobacco use within the past year (under-reporting)
Treatment Conditions	Initiation of Dapagliflozin vs. DPP-4i inhibitor	Initiation of Dapagliflozin vs. DPP-4i	Intention to initiate the study treatments (i.e., treatment allocation) will be

			emulated using the first observed prescription.
Endpoint	Time to first MACE (non-fatal MI, stroke, all-cause death)	Same: time to first MACE, defined using diagnostic codes in primary and secondary care and death registry data	Emulated using validated code lists
Summary Measure	Hazard Ratio	Hazard Ratio	
Intercurrent Events and Strategies to Handle Them	***Treatment discontinuation***: treatment policy ***Treatment switching***: treatment policy ***Addition of another antihyperglycemic agent:*** treatment policy ***All-cause death:*** composite strategy (included in endpoint)	Same: intercurrent events handled according to pre-specified strategies of the hypothetical target trial	

Rationale for handling of intercurrent events: Intercurrent events such as treatment discontinuation, treatment switching, or initiation of add-on antihyperglycemic therapies are handled using a treatment policy strategy in the Primary Estimand, as this approach aligns with regulatory expectations for estimating effects under usual clinical practice conditions and supports comparative effectiveness evaluations.

We include all-cause mortality in the composite strategy across estimands. This approach avoids differential handling of cardiovascular death (included in MACE) versus non-cardiovascular deaths (competing risk for MACE), since in RWD it is difficult to reliably distinguish between cardiovascular and non-cardiovascular causes of death.

6.2 Supplementary Estimand 2

Research questions answered by estimand 2: What is the HR of MACE for Dapa vs DPP-4i in patients with type 2 diabetes with or at risk for ASCVD while on treatment (i.e., before treatment discontinuation, switching or new add-on antihyperglycemic therapy)?

Table 3. Estimand 2

Attribute	Target Trial	Target Trial Emulation	Comment
Population	Same as table 2		

Treatment Conditions	Same as table 2		
Endpoint	Same as table 2		
Summary Measure	Same as table 2		
Intercurrent Events and Strategies to Handle Them	***Treatment discontinuation:*** while on treatment ***Treatment switching:*** while on treatment ***Addition of another antihyperglycemic agent***: while on treatment ***All-cause death:*** composite strategy (included in endpoint)	Same: intercurrent events handled according to pre-specified strategies of the hypothetical target trial Treatment discontinuation is measured using prescription refill data where a gap of more than 90 days is considered discontinuation Treatment switch is measured using prescription refill data in which discontinuation is defined as previously (a gap of more than 90 days in the sequence of prescriptions) and a switch is defined as the receipt of a new prescription for an anti-hyperglycaemic within this period. The index treatment discontinuation element is required to distinguish the addition of another antihyperglycemic agent (added on to the index therapy) from a treatment switch, which is defined as the initiation of another agent during continuous treatment with the index therapy. All cause death measured using death registry data	For while on treatment approach, mismeasurement of treatment discontinuation, switching or additional of another anti-hyperglycaemic events is an issue for the analysis since we are not interested in data after the occurrence of the IE. The accurate identification of these IEs depends on the assumptions made about prescription duration and allowed gaps between sequential prescriptions, which are used to identify the duration of the index treatment episode. These assumptions are evaluated in a sensitivity analysis described below.

Rationale for handling of intercurrent events: Uses a while-on-treatment strategy to describe the association between treatment exposure and outcomes during the period when patients are observed to remain adherent, providing complementary insights into biological treatment responsiveness. This approach does not by default yield a causal effect unless treatment discontinuation were ignorable, and it differs from the hypothetical strategy in that while-on-treatment relies only on observed data without imputing follow-up beyond treatment discontinuation.

6.3 Supplementary Estimand 3

Same as for estimand 1 (see table 2), except with RMST as the summary measure

Research question answered by estimand 3: What is the difference in restricted (at 3 and 5 years) mean survival time to MACE between Dapa and Dpp4-I in patients with type 2 diabetes with or at risk for ASCVD regardless of treatment discontinuation, switching or new add-on antihyperglycemic therapy?

7. Research methods

7.1 Study design

Research design (e.g. cohort, case-control, etc.): Active comparator new user; cohort study

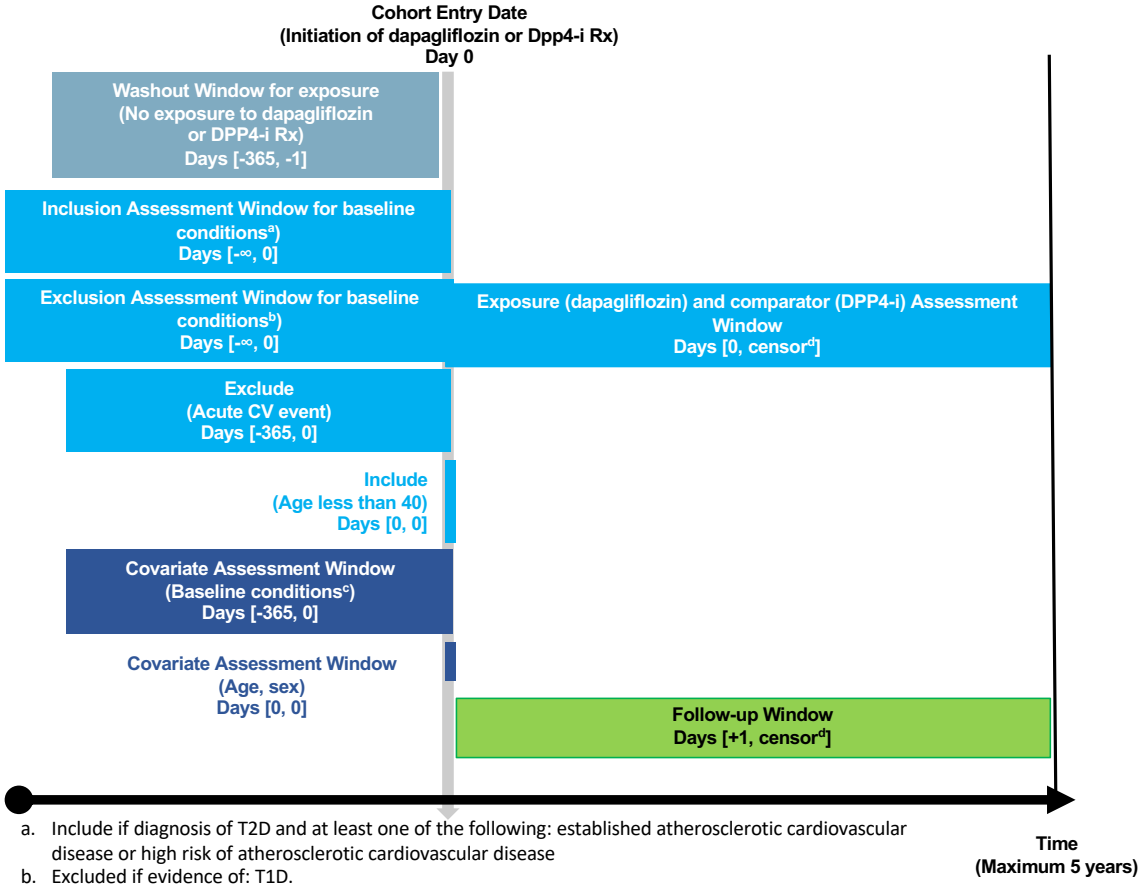
Study design choice:

Using an active comparator (rather than a non-user or general population) creates a more comparable reference group. DPP-4 inhibitors represent a clinically relevant and commonly prescribed alternative to SGLT2 inhibitors like dapagliflozin, and are considered cardiovascular neutral, which helps isolate the effect of dapagliflozin on cardiovascular outcomes.

By comparing new users of dapagliflozin to new users of DPP-4 inhibitors—both of which are guideline-recommended second-line treatments for T2D—the design helps ensure that patients initiating either therapy have similar indications and disease severity at baseline. This reduces confounding by disease severity or progression.

New user design aligns with the target trial for this specific study (in which the intervention is initiation of treatment after washout). The study aligns start of follow-up with treatment initiation, ensuring that all included patients are at risk of the outcome from a common time point. This avoids immortal time bias, which can occur if follow-up starts before the actual initiation of therapy.

7.2 Study design diagram



a. Include if diagnosis of T2D and at least one of the following: established atherosclerotic cardiovascular disease or high risk of atherosclerotic cardiovascular disease
b. Excluded if evidence of: T1D.
c. Full list and code algorithms provided in protocol
d. Censored at earliest of MACE, death, disenrollment or end of the study period (5 years after treatment initiation. Censoring is estimand specific (Censor after IE's for 2, do not censor afterwards for 1 & 3).

7.3 Setting

This study is conducted using routinely collected electronic health records from 2012 to 2024, reflecting the period of dapagliflozin use in routine clinical practice. The study is set primarily in primary care, drawing on longitudinal data from general practices with linkage to hospital data. Data are sourced from two European countries, the United Kingdom (Clinical Practice Research Datalink [CPRD]) and Spain (Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria [BIFAP]), providing population-based and representative coverage of real-world clinical care.

7.3.1 Definition of time 0 (and other primary time anchors) for entry to the study population

Time 0 (index date) is the date of initiation of dapagliflozin or DPP4 inhibitor. This is when patients enter the study population and mimics the timing of initiation of therapy at randomization in the target trial.

Table 4. Operational Definition of Time 0 (index date) and other primary time anchors

Study population name(s)	Time Anchor Description (e.g. time 0)	Number of entries	Type of entry	Washout window	Care Setting[1]	Code Type[2]	Diagnosis position	Incident with respect to…	Measurement characteristics/ validation	Source of algorithm
Exposure: dapagliflozin	Date of incident prescription for dapagliflozin (time 0)	Single	Incident	[-365, -1]	OP	ATC	n/a	Dapagliflozin or DPP4i (any formulation of either)		Investigator review of generic names
Comparator: DPP4i	Date of incident dispensation for DPP4i (time 0)	Single	Incident	[-365, -1]	OP	ATC	n/a	Dapagliflozin or DPP4i (any formulation of either)		Investigator review of generic names

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

[2]See appendix for listing of clinical codes for each study parameter

7.3.2 Study inclusion criteria:

Inclusion Criteria

1. Diagnosis of T2D prior to index
   Rationale: This ensures that all study participants have the condition of interest, type 2 diabetes, at the time they initiate treatment. It aligns the study population with the intended use of both dapagliflozin and DPP-4 inhibitors, which are approved for managing T2D.
2. Age 40 and over at index
   Rationale: Cardiovascular risk increases with age, and both clinical trial and real-world evidence suggest that the benefit-risk profile of glucose-lowering therapies may differ by age group. Including only patients aged ≥40 helps to target a population with a meaningful risk of major adverse cardiovascular events, thereby increasing the clinical relevance and event rate in the study.
3. Either established atherosclerotic cardiovascular disease (ACVD) or at high risk of ACVD.
   Rationale: Since the primary endpoint is MACE, the study aims to evaluate cardiovascular outcomes in individuals who are most likely to experience them. Including patients with established ACVD or those at high risk increases the baseline risk within the population for cardiovascular events, improving relevance to clinical decision-making.

Established ACVD defined as history or diagnosis of at least one of the following diagnosed anytime prior to the index date
Ischemic heart disease
Ischemic cerebrovascular disease
Peripheral arterial disease
Rationale: Established ACVD includes ischemic heart disease, ischemic cerebrovascular disease, or peripheral arterial disease, which are well-recognized manifestations of atherosclerosis and indicate existing cardiovascular burden.

High ACVD risk defined as no established ACVD, age ≥ 55 years in men and ≥ 60 in women and at least one of the following:
Documented history or diagnosis of dyslipidaemia anytime prior to index date
Current lipid lowering therapy at index
Documented history or diagnosis of hypertension anytime prior to index
Current anti-hypertensive medication use at index
Tobacco use within the last year or current use at index
Rationale: High ACVD risk is defined based on age and additional risk factors (e.g., dyslipidemia, hypertension, tobacco use), consistent with definitions used in major cardiovascular outcome trials like DECLARE-TIMI 58. This helps ensure that the population reflects real-world patients for whom cardiovascular risk reduction is a key treatment goal.

Table 5. Operational Definitions of Inclusion Criteria

Criterion	Details	Assessment window	Care Settings[1]	Code Type[2]	Diagnosis position[3]	Applied to study populations:	Measurement characteristics/validation	Source for algorithm
Type 2 Diabetes	Diagnosis of T2D prior to index	[-∞, 0]	OP, IP	ICD-10, SNOMED	Any	All study populations	In progress	In progress
Age	≥40 years at index date	[0, 0]	n/a	n/a	n/a	All study populations		
Established ACVD	Diagnosis of ischemic heart disease, ischemic cerebrovascular disease, or peripheral arterial disease	[-∞, 0]	OP, IP	ICD-10, SNOMED	Any	All study populations		
Ischemic heart disease	Includes diagnosis of myocardial infarction, angina, coronary artery disease	[-∞, 0]	OP, IP	ICD-10, SNOMED	Any	All study populations		
Ischemic cerebrovascular disease	Includes diagnosis of ischemic stroke, transient ischemic attack	[-∞, 0]	OP, IP	ICD-10, SNOMED	Any	All study populations		
Peripheral arterial disease	Includes diagnosis of claudication, arterial thrombosis, revascularization procedures	[-∞, 0]	OP, IP	ICD-10, SNOMED	Any	All study populations		
Dyslipidaemia	Diagnosis of dyslipidaemia	[-365, 0]	OP,IP	ICD-10, SNOMED	Any	All study populations		
Lipid lowering therapy	Active prescription[4] of lipid-lowering drug at index	[-30, 0]	OP	ATC	n/a	All study populations		
Hypertension	Diagnosis of hypertension	[-∞, 0]	OP	ICD-10, SNOMED	Any	All study populations		

Antihypertensive medication	Active prescription[4] at index date	[-30, 0]	OP	ATC	n/a	All study populations		
Tobacco use	Recorded tobacco use in past year or current use	[-365, 0]	OP	SNOMED	Any	All study populations		

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] See appendix for listing of clinical codes for each study parameter
[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)
[4] Active prescription refers to ongoing use at index, operationalised as a prescription within 30 days prior to index with coverage overlapping the index date.

7.3.3 Study exclusion criteria

Exclusion Criteria

Less than one year of recorded medical history in database

Rationale: ensures sufficient time to assess new use of the drugs of interest and adequate time for assessment of covariates at baseline and ensure all outcomes are incident

Treatment with SGLT2i or DPP-4i in the last year.

Rationale: ensures that the cohort includes new users of either dapagliflozin (SGLT2i) or DPP-4 inhibitors. By excluding individuals who have used either treatment class in the past year, the design emulates an incident user design, which aligns with the target trial and avoids biases associated with prior exposure (such as depletion of susceptibles or time-varying confounding by treatment history).

Documented history or diagnosis of acute cardiovascular event in the year prior to index

Rationale: This criterion ensures that all major adverse cardiovascular events (MACE) captured during follow-up are incident outcomes. Excluding patients with recent acute cardiovascular events ensures that exposure (treatment initiation) precedes the outcome, which is critical for establishing temporality.

Documented history or diagnosis of Type 1 diabetes at any time prior to index

Rationale: This study is focused specifically on the type 2 diabetes population. SGLT2i and DPP-4i are approved and indicated for T2D but not for type 1 diabetes. Including patients with T1D would mix treatment indications and risk profiles, potentially biasing results. This exclusion ensures that the study population reflects the regulatory and clinical indication for these treatments and ensures population homogeneity.

Table 6. Operational Definitions of Exclusion Criteria

Criterion	Details	Assessment window	Care Settings[1]	Code Type2	Diagnosis position3	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
SGLT2 Inhibitor Use	Any prescription for an SGLT2 inhibitor within the 365 days before index date	[-365, -1]	OP	ATC	n/a	All study populations	In progress	In progress
DPP-4 Inhibitor Use	Any prescription for a DPP-4 inhibitor within the 365 days before index date	[-365, -1]	OP	ATC	n/a	All study populations		
Acute Cardiovascular Event	Diagnosis of acute CV event (e.g., MI, unstable angina, stroke) in the year prior to index	[-365, -1]	IP, OP	ICD-10, SNOMED	Primary or secondary	All study populations		
Type 1 Diabetes	Diagnosis of type 1 diabetes at any point before index	[-∞, 0]	IP,OP	ICD-10, SNOMED	Primary or secondary	All study populations		

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] See appendix for listing of clinical codes for each study parameter
[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

7.4 Variables

7.4.1 Exposure(s) of interest

The exposure group consists of patients who initiate oral dapagliflozin at any dose, formulation, or regimen. This includes individuals receiving dapagliflozin as part of a combination therapy used as first-line treatment, as monotherapy for first-line treatment, or as a switch to or add-on therapy to any existing antihyperglycemic treatment. The intervention period for dapagliflozin may last for up to five years.

The comparator group consists of patients who initiate any oral DPP-4 inhibitor, at any dose, formulation, or regimen. This includes individuals receiving DPP-4 inhibitors as part of a combination therapy used as first-line treatment, as monotherapy for first-line treatment, or as a switch to or add-on therapy to any existing antihyperglycemic treatment. The intervention period for DPP-4 inhibitors may also last for up to five years.

The use of an active comparator is the most appropriate approach because it aligns with real-world clinical practice and helps to address confounding by indication. It allows for the safety and effectiveness of dapagliflozin to be benchmarked against another medication that is commonly prescribed to treat patients with type 2 diabetes. DPP-4 inhibitors are frequently used as an alternative to SGLT2 inhibitors in routine care. Furthermore, DPP-4 inhibitors are considered to have a neutral effect on cardiovascular outcomes, making them a suitable comparator for evaluating cardiovascular safety and effectiveness.

Algorithm to define duration of exposure effect:

For the construction of treatment episodes based on prescription records, the duration of each prescription is first derived using information available in the database. When end date or days supply is recorded, this information is used to directly define the prescription duration along with the date of the prescription. If days supply is not recorded, prescription duration will be estimated based on the quantity prescribed and dosing instructions. If dosing information is missing or incomplete, standard dosing assumptions consistent with routine clinical practice will be used (e.g., one defined daily dose according to WHO ATC classification). Overlapping days between prescriptions are handled by carrying forward any unused supply. Specifically, if a refill occurs before the end of the previous prescription's calculated days' supply, the overlapping days are added to the end of the new prescription's duration. We assume that each dispensed prescription has a lasting effect of up to 90 days. Therefore, a gap of up to 90 days between the end of a prescription's days' supply and the subsequent refill is allowed, without considering the patient as having discontinued treatment. The choice of a 90-day window aligns with common prescribing and dispensing practices for chronic medications such as glucose-lowering therapies and is consistent with prior pharmacoepidemiologic studies. This grace period accommodates typical variations in refill timing, medication stockpiling, or short treatment interruptions, thereby reducing the risk of misclassifying ongoing therapy as discontinued. Additionally, for the final prescription in a treatment episode, we extend exposure by 90 days beyond the calculated end date to account for any residual pharmacological effect or continued use.

Table 7. Operational Definitions of Exposure

Exposure group name(s)	Details	Washout window	Assessment Window	Care Setting[1]	Code Type[2]	Diagnosis position[3]	Applied to study populations:	Incident with respect to…	Measurement characteristics/ validation	Source of algorithm
Exposure	SGLT2-i	[-365, -1]	[0, censor]	OP	ATC	n/a	All study populations	Prescriptions measured in one year prior to first use	In progress	In progress
Comparator	DPP4-i	[-365, -1]	[0, censor]	OP	ATC	n/a	All study populations	Prescriptions measured in one year prior to first use		

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] See appendix for listing of clinical codes for each study parameter
[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter

7.4.2 Outcome(s) of interest

The primary endpoint will be time to the first occurrence of major adverse cardiovascular events (MACE). For this study a modified version of the MACE endpoint will be considered. MACE is a composite endpoint that includes myocardial infarction, ischemic stroke or cardiovascular death. In this study we will include all-cause death as part of the endpoint in order to implement a composite strategy to handle both cardiovascular and non-cardiovascular death. Components of MACE represent serious, life-threatening events that directly reflect the cardiovascular health of a patient**.** All-cause mortality is considered here, but death is expected to be mainly related to cerebrovascular, cardiovascular causes among those with type 2 diabetes

Table 8. Operational Definitions of Outcome

Outcome name	Details	Primary outcome?	Type of outcome	Washout window	Care Settings[1]	Code Type2	Diagnosis Position3	Applied to study populations:	Measurement characteristics/ validation	Source of algorithm
MACE (Composite outcome of AC Mortality/MI/Stroke)	Composite of all-cause mortality, myocardial infarction, and stroke	Yes	Time-to-event	[-365;0]	IP, OP, Mortality registry	ICD-10, SNOMED	Primary or secondary	Exposure: Dapagliflozin, Comparator: DPP4-I	In Progress	In Progress
Myocardial infarction	Non-fatal myocardial infarction identified from hospital discharge diagnoses. Hospitalization can also be recorded in Primary care records.	Yes (as part of composite MACE)	Time-to-event	[-365;0]	IP, OP	ICD-10, SNOMED	Primary or secondary	Exposure: Dapagliflozin, Comparator: DPP4-I		
Stroke	Non-fatal ischemic stroke identified from hospital discharge diagnoses. Hospitalization can also be recorded in Primary care records.	Yes (as part of composite MACE)	Time-to-event	[-365;0]	IP, OP	ICD-10, SNOMED	Primary or secondary	Exposure: Dapagliflozin, Comparator: DPP4-I		

All-cause mortality	Death from any cause, as recorded in mortality registry	Yes (as part of composite MACE)	Time-to-event	[-365;0]	Mortality registry	ICD-10	n/a	Exposure: Dapagliflozin, Comparator: DPP4-I		

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] See appendix for listing of clinical codes for each study parameter
[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

7.4.3 Follow-up

Table 9. Operational Definitions of Follow Up

Follow up start	Day 1	
Follow up end[1]	**Select all that apply**	**Specify**
Date of outcome	Yes	First occurrence of MACE
Date of death	Yes	Included as part of the composite outcome; AC death is included using a composite strategy
End of observation in data	Yes	Censor at date of last healthcare contact, known deregistration, or database end; non-administrative censoring
Day X following index date *(specify day)*	Yes	Day 1825 (5 years after index date); administrative censoring
End of study period (specify date)	Yes	31 December 2024
End of exposure *(specify operational details, e.g. stockpiling algorithm, grace period)*	Yes	Only used to define follow-up end for estimand 2 (where a "while on treatment" strategy is used to handle treatment discontinuation). Does not define end of follow-up for estimands 1 or 3.
Date of add to/switch from exposure *(specify algorithm)*	Yes	Only used to define follow-up end for estimand 2 (where a "while on treatment" strategy is used to handle treatment discontinuation). Does not define end of follow-up for estimands 1 or 3.

7.4.4 Covariates (confounding variables and effect modifiers, e.g. risk factors, comorbidities, comedications)

We identified demographics, comorbidities, comedications and other clinical variables that are risk factors for stroke, MI or all-cause mortality.

Table 10. Operational Definitions of Covariates

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type[2]	Diagnosis Position[3]	Measurement characteristics/validation	Source for algorithm
Demographics								
Age at index		Continuous	[0,0]			N/A		
Gender		Binary	[0,0]			N/A		
Frailty		Categorical (fit, mild, moderate, severe)	[-365,0]	IP, OP, OT	ICD-10, SNOMED, ATC	Any		
Duration of diabetes		Continuous	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED, ATC	Any		
HbA1c		Continuous	[-90,0]	OP, OT	, SNOMED	N/A		
Comorbidities	At least one recorded diagnosis in time window							
Obesity		Binary	[-365,0]	IP, OP, OT	ICD-10, SNOMED, ATC	Any		
Heart transplant		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED,	Any		
Microvascular complications		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED,	Any		
Severe hypoglycaemia		Binary	[-365,0]	IP, OP, OT	ICD-10, SNOMED, ATC, labs	Any		

Diabetic neuropathy		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED,	Any		
Diabetic Retinopathy		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED,	Any		
Keto-/lactate acidosis		Binary	[-365,0]	IP, OP, OT	ICD-10, SNOMED,	Any		
Lower limb amputations		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED,	Any		
COPD		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED, ATC	Any		
Cancer		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED, ATC	Any		
Major organ specific bleeding		Binary	[-365,0]	IP, OP, OT	ICD-10, SNOMED, ATC	Any		
Bariatric surgery		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED,	Any		
CKD stages 1-4		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED, ATC, labs	Any		
CKD stage 5		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED, ATC, labs	Any		
Kidney Transplantation		Binary	[-∞, 0]	IP, OP, OT	ICD-10, SNOMED,	Any		
Dialysis		Binary	[-365, 0]	IP, OP, OT	ICD-10, SNOMED,	Any		
Co-medications	At least one prescription in time window			OP, OT	ATC	N/A		
Metformin		Binary	[-365,0]	OP, OT	ATC			
Sulfonylurea		Binary	[-365,0]	OP, OT	ATC			
Sulfonamides		Binary	[-365,0]	OP, OT	ATC			
Alfa glucosidase inhibitors		Binary	[-365,0]	OP, OT	ATC			
Thiazolidinediones		Binary	[-365,0]	OP, OT	ATC			
Other GLDs		Binary	[-365,0]	OP, OT	ATC			

Insulin		Binary	[-365,0]	OP, OT	ATC			
ACE inhibitors		Binary	[-365,0]	OP, OT	ATC			
ARBs		Binary	[-365,0]	OP, OT	ATC			
Beta-blockers		Binary	[-365,0]	OP, OT	ATC			
Low/high ceiling diuretics		Binary	[-365,0]	OP, OT	ATC			
Aldosterone antagonists		Binary	[-365,0]	OP, OT	ATC			
Thiazide diuretics		Binary	[-365,0]	OP, OT	ATC			
Calcium channel blockers		Binary	[-365,0]	OP, OT	ATC			
Digitoxin/digoxin		Binary	[-365,0]	OP, OT	ATC			
Antiarrhythmics		Binary	[-365,0]	OP, OT	ATC			
Statins		Binary	[-365,0]	OP, OT	ATC			
Warfarin		Binary	[-365,0]	OP, OT	ATC			
Aspirin		Binary	[-365,0]	OP, OT	ATC			
Purinergic receptor $P2Y_{12}$ antagonists		Binary	[-365,0]	OP, OT	ATC			
Other antiplatelets		Binary	[-365,0]	OP, OT	ATC			
Corticosteroids		Binary	[-365,0]	OP, OT	ATC			
Weight loss drugs		Binary	[-365,0]	OP, OT	ATC			
Established ACVD				OP, IP	ICD10, SNOMED			
Ischemic heart disease			[-365,0]	IP, OP, OT	ICD-10, SNOMED, ATC			
Ischemic cerebrovascular disease			[-365,0]	IP, OP, OT	ICD-10, SNOMED, ATC			
Peripheral arterial disease			[-365,0]	IP, OP, OT	ICD-10, SNOMED, ATC			

ACVD Risk factors								
Dyslipidaemia			[-365,0]	IP, OP, OT	ICD-10, SNOMED, ATC			
Hypertension			[-365,0]	IP, OP, OT	ICD-10, SNOMED, ATC			
Tobacco use			[-365,0]	IP, OP, OT	SNOMED			

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] See appendix for listing of clinical codes for each study parameter
[3] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

7.5 Core Emulation Table – Design Summary

Table 11. Comparison of Target Trial and Proposed Target Trial Emulation Design Elements

Attribute	Target Trial	Target Trial Emulation	Comment
Eligibility	**Inclusion**: - Age ≥ 40 - Diagnosed with type 2 diabetes - Established ASCVD or high ASCVD risk (age ≥55 (men) or ≥60 (women) plus at least one cardiovascular risk factor(e.g., hypertension, dyslipidemia, tobacco use) **Exclusion**: - Prior use of SGLT2i or DPP-4i within the last year prior to randomisation - Acute cardiovascular event in past 12 months - Type 1 diabetes	**Inclusion:** Same: - Diagnosis codes for type 2 diabetes - Recorded ASCVD or ≥1 CV risk factors in baseline data - Treatment initiation with either dapa or Dpp-4i using prescription records All measured in the one year prior to the first prescription for either dapagliflozin or DPP4-i **Exclusion:** - Same: -Prior prescription of SGLT2i or DPP-4i based on prescription records - Type 1 diabetes identified from diagnostic codes	Emulation restricts to new users in routine care Eligibility applied using structured EHR data; may require proxy measures for ASCVD or risk factors Exposure and comorbidity definitions will be operationalized using prescription and diagnostic codes. We will apply lookback windows (e.g., one year to define incident use), while recognizing that accuracy may also depend on factors such as the choice of phenotyping algorithm, the placement of the lookback period, and the availability and reliability of underlying data.

		- Acute cardiovascular events measured using diagnostic codes – Medications are measured in the one year prior to the first prescription for either dapagliflozin or DPP4-I; Chronic conditions are measured at any point prior to this index date.	
Setting	Multicentre	Recruitment of patients for a multicentre study will be emulated by selecting patients who are seen in several primary care clinics	Reflects the setting from which patients are most likely to be recruited from. Will be missing hospital setting for recruitment, but T2DM patients are most likely to be managed in primary care. Although measurement of characteristics (comorbidities) can be conducted using both inpatient and outpatient information, the study setting still reflects those seen in primary care since the represents the base study population in RWD sources.
Treatment conditions	Dapagliflozin and DPP-4 inhibitors, each potentially added to usual care, reflecting real-world use without restrictions on dose or treatment duration.	Initiation of dapagliflozin or DPP-4i measured using first prescription of each medication. Added to usual care, meaning in addition to any other antihyperglycemic therapy the patient may already be prescribed.	Dose or duration flexibility mirrors routine care, as does being added to background therapy, although this introduces some uncertainty since the intervention may take several forms. However, these variations can be considered largely exchangeable within the treatment strategies. Potential mismeasurement of treatment initiation may also occur due to non-adherence.
Treatment Assignment	Simple 1:1 randomisation	Assignment reflects clinical need. Inverse probability of treatment weighting (IPTW) will be used to adjust for baseline confounders.	Randomisation cannot be directly emulated. IPTW will be used in the statistical analysis to balance confounders in absence of randomisation.
Follow-up	Begins at randomisation; ends at first occurrence of outcome, study withdrawal, loss to follow-up, or at 5 years after randomisation	Begins at treatment initiation which is first prescription of dapagliflozin or DPP-4i; ends at outcome, loss to follow-up or at 5 years after treatment initiation.	Aligns start of follow-up with treatment initiation to mimic start of trial; loss to follow-up can be identified in data sources as de-registration from general practices, migration
Outcome	Time to first MACE: composite of non-fatal MI, stroke, CV or non-CV death	Same composite outcome identified using diagnostic and mortality records in linked databases	Code lists and outcome definitions validated or informed by prior CVOT emulations

Intercurrent Events and Strategies to Handle Them	***Treatment discontinuation***: treatment policy ***Treatment switching***: treatment policy ***Addition of another antihyperglycemic agent:*** treatment policy ***All-cause death:*** composite strategy (included in endpoint)	Same but measured based on prescribing data and mortality data	Accurately identifying treatment discontinuation and switching will not be a limitation for estimand 1 since they are ignored under the treatment policy approach
Loss to follow-up	Patients who fail to return for the required study visits and their health condition and vital status remains unknown despite multiple attempts to contact them.	Patients with known deregistration date, practice withdrawal or database end. This is directly measured in RWD source.	Loss to follow-up will be defined using real-world proxies, recognizing that in some cases patients may appear to remain under follow-up despite having effectively left (e.g., if they do not formally de-register from their GP). This risk is expected to be low, where unique patient identifiers ensure automatic de-registration upon re-registration at a new practice.

Estimand 2

For estimand 2, accurately identifying treatment discontinuation and switching will be a limitation since the while on treatment approach is used and as exposure duration and time to switch may be underestimated.

Operational definitions include

- Treatment discontinuation is identified using prescription refill data, where a gap of more than 90 days between refills is considered a discontinuation.
- Treatment switching is similarly measured using prescription records, with a gap of more than 90 days and receipt of a new antihyperglycemic indicating a switch to a new therapy.
- All cause death determined using death registry data

Estimand 3

Accurately identifying treatment discontinuation and switching will be a not be a limitation since they are ignored under the treatment policy approach

7.6 Data analysis

7.6.1 Analysis Plan

Overview

The analyses are conducted within a target trial emulation framework to estimate the effect of dapagliflozin compared with DPP-4 inhibitors on the risk of major adverse cardiovascular events (MACE).

For **Estimand 1**, the main estimand supporting decision making, the primary causal effect summary measure is the hazard ratio for time to first MACE, estimated using an inverse probability of treatment weighted (IPTW) Cox proportional hazards model. The Cox model will be fitted separately within each data source (CPRD and BIFAP), and the resulting hazard ratios will be combined using a random-effects meta-analysis; potential sources of heterogeneity will be described qualitatively, including structural differences (e.g., coding systems, population coverage) and measurement differences (e.g., recording practices) and their implications (e.g., residual confounding or misclassification).

Sensitivity analyses will assess robustness of the primary findings to key assumptions, including inverse probability of censoring weighting (IPCW), tipping point analysis, and probabilistic bias analysis for non-differential exposure misclassification (details in Section 7.6.5).

Two supplemental estimands are also defined: **Estimand 2**, applying a while-on-treatment strategy for intercurrent events, and **Estimand 3**, estimating treatment effects using restricted mean survival time (RMST) derived from an IPTW-weighted Weibull accelerated failure time (AFT) model. In addition, supplemental analyses (e.g., crude and IPTW-adjusted Kaplan–Meier curves, crude Cox models, event counts and incidence rates, propensity score and weight distributions, covariate balance before and after weighting, censoring and intercurrent event patterns, proportional hazards diagnostics, positivity checks, and multiple-imputation diagnostics) will be conducted to support interpretation of the main analysis.

7.6.2 Primary Estimand (1) Analysis

i. Objective

Estimate the effect of dapagliflozin on the risk of MACE compared to DPP-4 inhibitors.

ii. Exposure contrast

Dapagliflozin vs DPP-4 inhibitors.

iii. Outcome

Time to first major adverse cardiovascular event (composite endpoint including stroke, myocardial infarction, and all-cause death).

iv. Analytic software:

R.

v. Handling of intercurrent events

Intercurrent events are handled as follows:

- Treatment discontinuation: treatment policy strategy
- Treatment switching: treatment policy strategy
- Addition of another antihyperglycemic agent: treatment policy strategy

For the intercurrent events handled using treatment policy approach, observed follow-up after intercurrent events is retained because follow-up data after intercurrent events are of interest.

- All-cause death: composite strategy by including all-cause death as part of the primary composite endpoint

vi. Outcome Modelling

A Cox proportional hazards model, weighted by inverse probability of treatment (IPTW), will be used to estimate the effect of initiating dapagliflozin versus a DPP-4 inhibitor on time to first major adverse cardiovascular event (MACE).

- Start of follow-up: date of treatment initiation for either dapagliflozin or DPP4-i
- Timescale: time since index
- Endpoint: first occurrence of MACE.
- Censoring:
    - Non-administrative censoring: Loss to follow-up
    - Administrative censoring: End of study follow-up in the absence of MACE (maximum 5-years)
- Model covariate: Treatment group (dapagliflozin vs. DPP-4i)

Assumptions of KM

- *Non-informative Censoring:*
    - For Crude KM curves: Censoring is assumed to be independent of the outcome, ***conditional*** on treatment and survival up to the time of censoring.
- For IPTW adjusted KM: *Censoring is assumed to be independent of the outcome, conditional on the treatment, survival up to the time of censoring and indirectly baseline covariates used to estimate the treatment weights*
- Times of events and censoring are known and correctly recorded
- Independence between observations

Assumptions of Cox Model

Proportional Hazards:

- The hazard ratio is assumed to be constant over time.

Non-informative Censoring:

- *Censoring is assumed to be independent of the outcome, conditional on the treatment, survival up to the time of censoring and indirectly baseline covariates used to estimate the treatment weights*

- Times of events and censoring are known and correctly recorded
- Independence between observations

Diagnostics for Cox Model

Proportional hazards assessed using log(-log) survival plots or Schoenfeld residuals

vii. Confounding Adjustment

Inverse Probability of Treatment Weighting (IPTW)

To adjust for baseline confounding, inverse probability of treatment weighting will be used. Propensity scores, defined as the probability of initiating dapagliflozin versus a DPP-4 inhibitor, will be estimated using logistic regression. The model will include baseline covariates selected a priori based on clinical relevance and prior evidence as potential confounders—specifically, variables considered to be associated with the outcome and plausibly associated with treatment but not affected by treatment initiation. These include patient demographics, lifestyle factors, comedications, comorbidities, and laboratory measures (e.g., HbA1c). Variables associated with the treatment alone or variables which are a common effect of the exposure and another variable that influences the outcome will be excluded. A directed acyclic graph (DAG) will inform the conceptual identification of confounding structures, although it is not formally presented here. Restricted cubic splines will be used to model continuous variables (e.g., age, diabetes duration).

Stabilized weights will be calculated by dividing the marginal probability of receiving the treatment actually received (i.e., the overall proportion treated in the study population) by the individual's estimated propensity score (i.e., the conditional probability of receiving their observed treatment). Weights will be truncated at the 1st and 99th percentiles to limit the influence of extreme values.

Weight truncation reduces the influence of individuals with highly improbable treatment assignments but does not resolve propensity score non-overlap. Therefore, if regions of the propensity score distribution show insufficient overlap, we plan to restrict analyses to the overlapping region (trimming) or apply overlap weights.

Truncated stabilized IPTW weights will then be applied in the Cox proportional hazards model (weighted likelihood) to estimate the marginal treatment effect (dapa vs dpp4i) on time to first MACE. Standard errors will be estimated using robust (sandwich) variance estimators.

Assumptions Underlying IPTW

- No unmeasured confounding (all relevant baseline confounders are included in the propensity score model).
- Positivity (each individual has a non-zero probability of receiving either treatment, given their covariates).
- Correct model specification (the propensity score model is correctly specified [functional form, covariate inclusion]).
- Consistency (each individual's potential outcome under the observed treatment equals their actual outcome).

Diagnostics for IPTW

- Covariate balance: Check that baseline characteristics are balanced across treatment groups after weighting.
    - Evaluate standardized mean differences (SMDs): SMDs < 0.1 will be considered acceptable.
- Positivity check: Evaluate the overlap in propensity score distributions between treatment groups (graphically).
    - Figures representing the distribution of IPTW for each treatment arm (to be conducted before and after truncation)

viii. Missing Data handling

Missing Exposure Data

We assume that missing refill or prescription records for dapagliflozin or DPP-4 inhibitors reflect true treatment discontinuation after 90 days, and not incomplete data capture or prescriptions issued outside the database

Missing Outcome Data

The Cox proportional hazards model implicitly assumes non-informative censoring, meaning that censored participants contribute time at risk up to the time of censoring and their censoring is unrelated to the outcome, conditional on model covariates (the same covariates used to estimate the inverse probability of treatment weights) and survival up to the time of censoring (i.e., outcome data is missing at random under these assumptions). In addition, IPTW may partially mitigate bias due to informative censoring insofar as censoring depends on these measured covariates.

It is also assumed that participants who do not experience the outcome before censoring are correctly classified as having not had the event—that is, the available data provide complete outcome coverage with respect to the defined endpoint. This assumption concerns the correct classification of the outcome variable, and violations (e.g., missed or delayed event recording) could lead to outcome misclassification.

Missing Covariate Data

The absence of a disease diagnosis code will be interpreted as the absence of the corresponding condition. Missing values for lifestyle factors (e.g., smoking status) and lab variables (e.g, HbA1c to measure hypoglycemia and creatinine to measure eGFR for CKD) will be addressed using multiple imputation with chained equations (MICE) under the Missing at Random (MAR) assumption.

Assessment of Missingness

Before performing imputation, we will examine the extent and patterns of missingness to evaluate whether imputation is appropriate. Specifically, we will:

- Quantify the percentage of missing data for each covariate.
- Compare the proportion of missing values across treatment groups to assess differential missingness.
- If a covariate has more than 40% missing data, we will consider alternative approaches (e.g., exclusion of the variable, sensitivity analyses) and justify the decision. Thresholds of 40% have been cited because effect estimates begin to be less reliable as the level of missingness increases beyond this threshold.[Jakobsen et al. (2017), “When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts”]

Imputation Model

The MICE procedure will include all covariates used in the outcome and treatment models, as well as predictors of missingness (if there are any additional factors not covered by the covariates in the treatment and outcome models). The treatment and outcomes of interest will also be included.

Key covariates included in the imputation model will be:

- Demographics (age, sex)
- Clinical history and comorbidities (e.g., cardiovascular disease, CKD, hypoglycemia, obesity)
- Laboratory values and vital signs (e.g., BMI, blood pressure, HbA1c, eGFR)
- Lifestyle factors (e.g., smoking)
- Medication use (e.g., antihypertensives, glucose-lowering drugs, statins, anticoagulants, antiplatelet agents, corticosteroids, etc.)

Full Conditional Distributions

MICE will use variable-specific conditional models:

- Logistic regression for binary variables (e.g., smoking yes/no).
- Multinomial logistic regression for categorical variables with >2 categories.
- Predictive mean matching for continuous variables

Number of Imputations and Diagnostics
We will generate at least 10 imputed datasets (to ensure stable estimates given the level of missingness) and pool results across imputations using Rubin's rules. Diagnostics will include:

- Checking whether imputed values are plausible and consistent with observed distributions.
- Evaluating convergence of the chained equations.
- Assessing stability and consistency of results across imputed datasets.

Effect Estimation Under Multiple Imputation
The imputation model will be applied prior to IPTW and effect estimation. IPTW and outcome models will then be fitted in each imputed dataset, producing treatment effect estimates (e.g., hazard ratios) and corresponding variances. These estimates will be combined across the imputed datasets using Rubin's rules, which account for both within-imputation variance (the average estimation error within each imputed dataset) and between-imputation variance (the variability in estimates across imputations). The total variance therefore reflects uncertainty from both the imputation process and the effect estimation, producing valid confidence intervals.

7.6.3 Supplemental Estimand (2) Analysis

Same as primary estimand but intercurrent events are handled using the following:

- Treatment discontinuation: Apply a while on treatment strategy
- Treatment switching: Apply a while on treatment strategy
- Addition of another antihyperglycemic agent: Apply a while on treatment strategy

Administrative censoring occurs at the occurrence of each IE because outcome data after the IE are no longer of interest. Therefore, there is no "missing outcome" data.

7.6.4 Supplemental Estimand (3) Analysis

Same as 7.5.2, except outcome model is IPTW weighted Accelerated Failure Time (AFT) model, followed by estimation of Restricted Mean Survival Time (RMST) at fixed time points (3 and 5 years).

Model Assumptions:

- Survival times follow a Weibull distribution
- Non-informative censoring (conditional on included covariates and survival up to time t)

– Log-linear relationship between covariates and log survival time

To estimate the RMST at 3 and 5 years from the Weibull AFT model, we first use the model to obtain the predicted survival curve for each treatment group. The RMST is then calculated as the average survival time up to a fixed time point, which corresponds to the area under the survival curve between time zero and the chosen time horizon (3 or 5 years).

- Fit the Weibull AFT model, which gives the shape and scale of the survival curve for each group.
- From this model, generate the predicted survival probability at each time primary
- Integrate (i.e., add up) the survival probabilities from time 0 to 3 years and separately from time 0 to 5 years. The result is the expected survival time lived within those windows.
- Compare the RMST values between treatment groups to obtain the difference in average survival time over 3 and 5 years.

7.6.5 Sensitivity Analyses

Table 12. Sensitivity analyses -Inverse Probability of Censoring Weighting (IPCW)

Analysis Methods	This analysis will examine the impact of varying assumptions about the censoring-at-random condition on the estimated treatment effect. In the primary analysis, we assumed that censoring is independent of the outcome, conditional on the treatment, survival up to the time of censoring and indirectly baseline covariates used to estimate the treatment weights. Inverse probability of treatment weights were used to reweight the sample so that, in the weighted pseudo-population, treatment assignment is independent of measured covariates. In the IPCW analysis, inverse probability of censoring weights will be additionally applied to reweight observations. It assumes censoring is independent of the outcome, with all common causes of both the outcome and censoring being directly accounted for. Follow-up will be divided into equal 30-day intervals. At the start of each interval, we will update the information available on each patient and assess whether they remain followed or have been censored. If they remain under observation, they contribute to the next risk set for that interval. A separate conditional probability of remaining uncensored is estimated for each interval and a time-specific weight component is calculated as: 1/(the estimated probability of remaining uncensored, given a set of baseline and time-updated covariates that could affect both censoring and the outcome). Characteristics that could affect both censoring and the outcome include: treatment group (dapa vs dpp4-i), demographics (age, sex), comedications (e.g., antihypertensives) and comorbidities (e.g., CKD)

	These time-specific weight components are then multiplied across intervals to form the cumulative weight up to time. The denominator probability will be estimated using pooled logistic regression fit to the person-interval dataset. In this model, the outcome is whether the participant was censored in that interval. Time will be modelled flexibly using restricted cubic splines. We will truncate weights at prespecified percentiles (1st and 99th). The IPCW will be applied as time-varying weights in the Cox model for time to first MACE. Because inverse probability of treatment weights (IPTW) are also used, the final analysis weights will be the product of IPTW and IPCW.
Assumptions	- Censoring is **conditionally** independent of the outcome given covariates (i.e., non-informative censoring/censoring at random, conditional on additional covariates beyond those in the outcome model). - Correct model specification - Positivity (every individual has a non-zero probability of remaining uncensored at each time point given their covariates) - Outcome does not directly influence its own missingness (would imply informative censoring via MNAR mechanism)
What is Being Varied?	- The condition of the censoring at random assumption. - Tests a different missing-at-random (MAR) assumption
Why (Objective)	- To evaluate whether the treatment effect estimate is sensitive to changes in the condition of the missing at random assumption
Strengths	- IPCW adjusts for common causes (measured) of censoring and the outcome -
Limitations	- IPCW is sensitive to model misspecification. - Cannot account for unmeasured factors affecting censoring. - Weighting can increase variance, especially if weights are unstable - Sensitivity analyses rely on varying the assumptions of the primary analysis, but for censoring these assumptions cannot be verified from the observed data; their plausibility can be discussed yet ultimately remains unknown.

Table 13. Sensitivity analyses -Tipping Point Analysis

Analysis Methods	Tipping point analysis is used to examine how the estimated treatment effect changes when the censoring mechanism deviates from the primary analysis assumption of censoring at random (CAR), conditional on treatment group and survival time. To implement the tipping point analysis, we will vary assumptions on the hazard rate after the censoring time for subjects who are non-administratively censored and impute a time-to-event to those participants under the assumed hazard rates. Assumptions about the possible risk of outcome post-censoring will be explored by varying the hazard of the event after censoring over a continuum of plausible values, rather than only assuming extreme scenarios (e.g., that all censored individuals immediately experience the event or that none do). Specifically, we will modify the estimated baseline hazard by applying a range of multipliers to reflect increasingly favorable to increasingly unfavorable censoring assumptions. The baseline hazard is estimated from the observed, uncensored follow-up under the Cox model used in the primary analysis. After censoring, this baseline hazard is modified using multiplicative sensitivity parameters to define post-censoring event-time distributions. For each scenario, post-censoring event times will then be repeatedly imputed by drawing from the resulting distributions, so that imputation reflects uncertainty in event timing rather than assigning deterministic extreme values. These assumptions will be varied independently in each treatment group to allow for potentially differential censoring mechanisms. For each combination of assumed post-censoring hazard rates in the two treatments, multiple imputation of the time-to-event will be performed. The primary analysis will then be performed for each imputed dataset with complete data (observed or imputed) for all subjects, and the resulting treatment effects should be combined across imputations. For each grid point we will then obtain a treatment effect estimate and associated uncertainty (95% confidence interval) which will allow us to determine what pattern of assumed post-censoring changes in hazard would have led to different conclusions. For example the point at which a treatment effect would no longer be distinguishable from a chance finding, or vice versa. The plausibility of the missingness pattern underpinning the tipping point will then be evaluated.
Assumptions	- No single assumed model. - Does not require formal modelling of the censoring mechanism but does require modelling the hazard of event post-censoring.
What is Being Varied?	- Tipping point analysis explores what assumptions about censored outcomes would be needed to be made to alter conclusions with respect to the treatment effect
Why (Objective)	- To evaluate whether the treatment effect estimate is sensitive to violations of the non-informative censoring assumption made in the primary analysis. - If the results remain consistent in direction and statistical significance across plausible censoring scenarios, this increases confidence that the primary findings are sensitive to the assumption made in the primary analysis.

Strengths	- Tipping point analysis does not require modelling assumptions.
Limitations	

Table 14. Sensitivity analyses -Non-Differential Exposure Misclassification

Analysis Method	***Probabilistic Bias Analysis using Monte Carlo Simulation at the summary estimate measure*** To evaluate the potential impact of exposure initiation misclassification at index date, we will conduct a probabilistic bias analysis at the summary estimate level. This analysis will be performed after pooling hazard ratios across imputed datasets using Rubin's rules so that uncertainty due to missing data and sampling variability is incorporated into the input effect estimate. The pooled hazard ratio and its variance will then serve as the starting point for the bias analysis. Plausible distributions for the sensitivity and specificity of prescription‐based exposure classification will be specified based on the known limitations of prescription records as proxies for true medication use (e.g., prescriptions may not be filled or taken). This sensitivity analysis addresses misclassification of actual treatment initiation as inferred from prescription records (i.e. whether a recorded prescription corresponds to true initiation of use), and does not address misclassification of treatment discontinuation or adherence during follow-up. Within each Monte Carlo iteration, a hazard ratio will be sampled from a probability distribution informed by the pooled hazard ratio and its variance from the main analysis. This step propagates uncertainty due to sampling variability and multiple imputation into the bias analysis A single pair of sensitivity and specificity values will be drawn from the prespecified distributions and applied to both treatment groups, reflecting the assumption of non-differential exposure misclassification. This approach avoids introducing artificial random differences in initiation of medication use between treatment arms while allowing uncertainty in the overall degree of misclassification to be propagated. These values will be used to probabilistically adjust the pooled hazard ratio for exposure misclassification using established bias‐adjustment methods for dichotomous exposure classification[Banack et al. Monte Carlo Simulation Approaches for Quantitative Bias Analysis: A Tutorial. Epidemiol Rev. 2022; Fox et al. SAS and R code for probabilistic quantitative bias analysis for misclassified binary variables and binary unmeasured confounders, International Journal of Epidemiology 2023] . Repeating this process across many iterations (e.g., 10,000) will generate a distribution of bias‐adjusted hazard ratios that incorporates both (1) the uncertainty from the primary analysis (due to sampling variation and multiple imputation) and (2) the uncertainty associated with assumptions about misclassification parameters. Final uncertainty will be represented by the percentile‐based 95% confidence interval from the distribution of simulated bias‐adjusted hazard ratios, which is derived from the empirical distribution of bias-adjusted point estimates across all Monte Carlo iterations and incorporates uncertainty from sampling variability, multiple imputation, and exposure misclassification, rather than from confidence intervals calculated within individual iterations.

	Plausible parameter ranges for prescription-based exposure measurement in real-world data:[8] Sensitivity: 0.70-0.90 o lower bound reflects non-initiation after prescribing; upper bound reflects high adherence Specificity: 0.95-0.99 o high specificity expected, as those without prescriptions are unlikely to be truly exposed There is no reason to suspect that misclassification of actual prescription initiation is differential with respect to the outcome, since prescribing occurs before the outcome; any association between prescribing and the outcome is more plausibly explained by common causes, which are addressed through IPTW (in which prescribing would be independent of measured outcome determinants).
Assumptions	- Non-differential misclassification: exposure error is unrelated to the outcome. - Sensitivity and Specificity are correctly specified and assumed to apply uniformly at the population level (i.e. not varying across individuals). - The primary model (i.e. the IPTW-weighted Cox proportional hazards model used to estimate the treatment effect) is correctly specified.
What is Being Varied?	- The assumption of accurate treatment initiation measurement via prescribing records - Simulate variation in Sensitivity and Specificity to generate corrected exposure measurement. - Assess how misclassification could bias the estimated hazard ratio.
Why (Objective)	- Evaluate whether the findings are robust to exposure misclassification. - Determine whether plausible levels of misclassification could meaningfully alter the estimated treatment effect.
Strengths	- Explicitly incorporates uncertainty in exposure classification. - Produces a distribution of corrected effect estimates, not just a point estimate.
Limitations	- Requires accurate or defensible assumptions about Se and Sp. - Limited to non-differential misclassification unless extended. - Does not address misclassification of covariates or outcomes.

7.6.6 Other Supplemental Analyses

Baseline characteristics will be presented overall and stratified by treatment group. Categorical and binary variables will be summarized as counts (n) and percentages, while continuous variables will be reported using means and standard deviations or medians and interquartile ranges, as appropriate.

Kaplan–Meier methods will be used to compare the time-to-event distribution of MACE between patients treated with dapagliflozin and those treated with DPP-4 inhibitors. Crude Kaplan–Meier cumulative incidence curves will be estimated separately for patients initiating dapagliflozin and for those initiating a DPP-4 inhibitor. Time will be measured from the day after treatment initiation (index date) until the first occurrence of MACE or censoring.

The cumulative incidence (absolute risk) of MACE will be estimated using Kaplan–Meier methods over the entire follow-up period for each treatment group and presented graphically as full cumulative incidence curves. In addition, cumulative incidence will be reported at the pre-specified time points of 3 and 5 years, together with 95% confidence intervals. Values at 3 and 5 years will be obtained by evaluating the Kaplan–Meier step function at those time points.

Inverse probability of treatment weighted (IPTW) Kaplan–Meier curves will also be estimated.

We will also conduct descriptive analyses to characterize censoring patterns overall and across treatment groups. This will include median (IQR) time to censoring overall and according to the reason for censoring. This will be estimated separately for the overall study population and by treatment arm (dapagliflozin vs. DPP-4 inhibitors).

Reasons for censoring will include:
Administrative censoring: reaching the maximum follow-up period of 5 years or the end of the study period (31 December 2024).
End of data availability: last recorded healthcare encounter, database end date, or practice withdrawal.
Loss to follow-up: deregistration from the contributing practice or migration out of the healthcare system.

We will also compare baseline characteristics between eligible population and resulting study population after PS based trimming

7.6.7 Core Emulation Tables – Estimation Summary

Table 15. Core Emulation Table Estimand 1: Estimation Summary

Attribute	Target Trial	Target Trial Emulation	Comment
Analysis Method	Cox proportional hazards model to estimate HR for time to first MACE. Randomization ensures balance in measured and unmeasured confounders	Cox proportional hazards model weighted by stabilized IPTW, estimated separately in each data source (CPRD and BIFAP); pooled using random-effects meta-analysis. The analysis is conducted in in the trimmed population	IPTW used to emulate randomization in observational data Trimming of observations based on PS distribution represents a departure from the original target trial but is considered best practice when using propensity score methods in emulation. By removing patients in regions of non-overlap, the analysis is restricted to a population where treatment assignment is more comparable across groups. As a result, the estimated effect no longer applies to the entire eligible population but to this more comparable subset. If some patients have an extremely low probability of receiving one of the treatments, valid causal contrasts cannot be identified for them. Without trimming, effect estimates in these regions rely on unsupported extrapolation, making the results unstable and potentially biased.
Missing Data Assumptions and Methods	**Outcome**: Assumes non-informative censoring conditional on treatment, and survival time; censored participants contribute partial information. **Exposure**: N/A (trial monitoring ensures exposure data completeness) **Covariates**: Minimized through trial data collection	**Outcome**: same **Exposure:** For missing exposure data, assume absence of refill or prescription records for dapagliflozin or DPP-4 inhibitors indicates true treatment discontinuation after 90 days. **Covariates:** absence of a diagnosis code will be interpreted as absence of the condition, while missing lifestyle and laboratory variables will be imputed using multiple imputation by chained equations (MICE) under the missing at random assumption.	Mechanisms of missing exposure, covariate and outcome data differs between target trial and emulation (e.g., rather than leaving study, patients could be part of GP practice that no longer contributes data). Missing exposure data not possible in target trial but could be as a result of missing or incomplete prescription records in emulation. Multiple imputation would not occur for missing covariate data in target trial.

Statistical Model Assumptions	Proportional hazards assumption for Cox model. Censoring is non-informative (given assumption re: missing outcome data)	Same; proportional hazards assumption assessed with Schoenfeld residuals and log(-log) plots. **IPTW Assumptions**: no unmeasured confounding, positivity, correct model specification, consistency	Due to non-random treatment allocation, additional assumptions are required to identify the effect of treatment. Some of these assumptions difficult to verify (e.g., unmeasured confounding). Can argue consistency may be violated as a result of allowing variables doses and medications as part of treatment arm. Correct model specification checked by evaluating SMD in baseline characteristics after weighting.
Sensitivity Analyses	None	**IPCW:** Varies conditions of the censoring at random assumption (changed to include covariates (baseline and time-updated). **Tipping Point Analysis:** does not make assumption about the censoring mechanism, thus it is equally valid under MNAR or MAR **Probabilistic Bias Analysis:** Monte Carlo simulation to assess impact of non-differential exposure misclassification	Potential for exposure mismeasurement only present in emulation since exposure based on prescription records and assume adherence to prescribed treatment.

Estimand 2: Same as estimand 1 but without sensitivity analyses

Table 16. Core Emulation Table Estimand 3: Estimation Summary

Attribute	Target Trial	Target Trial Emulation	Comment
Analysis Method	Accelerated failure time (AFT) model assuming Weibull distribution	IPT weighted Weibull AFT model, followed by estimation of restricted mean survival time (RMST) at years 3 and 5	Same as table 1
Missing Data Assumptions and Methods to Handle	Same as table 1		
Statistical Model Assumptions	Weibull survival distribution; log-linear relationship between covariates and log survival time	Same; assessed using diagnostics such as log(-log(S(t))) vs log(t) for Weibull assumption and Q-Q plot for residuals	
Sensitivity Analyses	N/A		

7.7 Data sources

7.7.1 Data sources and Quality

Reason for selection / Rationale for selection and feasibility:
CPRD (GOLD and Aurum) and BIFAP were selected as data sources for this study because they offer large, high-quality, population-based electronic health records in the UK and Spain, respectively. Both sources provide the required data elements to operationalize the study design, including demographics, diagnoses, prescriptions, laboratory test results, hospitalizations, and mortality data. BIFAP allows for linkage to hospital discharge and mortality registries in several autonomous communities, while CPRD offers linkage to hospital and mortality records through NHS Digital. Both databases have previously been used in studies addressing cardiovascular outcomes and adverse drug effects and are known to support the emulation of clinical trial eligibility criteria and endpoints. Their extensive follow-up time (median ~10 years for BIFAP and ~6–13 years for CPRD) makes them suitable for assessing long-term outcomes. Data extraction and feasibility assessments confirm that the study variables are available with sufficient completeness and temporal coverage to support the research objectives.

Strengths of data source(s):
BIFAP provides high completeness and reliability for critical variables such as date of birth, sex, diagnostic codes (SNOMED, ICD-9/10), and prescription data (ATC codes). Hospital discharge data and mortality data are available for a large proportion of patients. Its population includes over 8 million individuals, with a median follow-up of 10 years. Similarly, CPRD covers over 19 million patients in total, with 16.5 million current acceptable patients in Aurum and nearly 3 million in GOLD. The data are updated frequently (monthly for GOLD, quarterly for Aurum), and mortality, prescribing, and diagnosis information are available with high completeness. CPRD includes validated linkages to secondary care, cancer registries, and national death registries. Both databases support validated outcome definitions and have been widely used in pharmacoepidemiological studies.u

Limitations of data source(s) (with potential impact on study results):
For BIFAP, smoking status is missing in approximately 56% of diabetic patients, which may impact adjustment for cardiovascular risk. Some variables (e.g., frailty, BMI, lab values) may have inconsistent completeness or be missing entirely for subpopulations. Hospital linkage is only available in some regions, which may limit outcome ascertainment for hospital-only diagnoses. Cause of death is only available for a subset of the population and may be misclassified.

For CPRD, only the year of birth is available for adults, which may slightly impact precision in age-based eligibility criteria. Ethnicity is missing for approximately 20% of patients overall (higher in earlier years). While vaccine and prescription data are complete, medication adherence is not directly observable. Linked hospital and mortality data are available only for practices in England and may be delayed.

Data Quality:

BIFAP is curated and maintained by the Spanish Agency for Medicines and Medical Devices (AEMPS). Data are pseudonymized, extracted from autonomous communities through secure pipelines, and cleaned and standardized centrally. SOPs for extraction, data curation, and quality control are available online:

CPRD is managed by the Medicines and Healthcare products Regulatory Agency (MHRA) in the UK. Data from participating general practices are extracted monthly (GOLD) or quarterly (Aurum) and undergo multi-level validation and quality checks. The CPRD provides data through its secure Trusted Research Environment or via multi-study licenses. Full metadata and SOPs are available at https://www.cprd.com/data-access and https://www.cprd.com/data-quality

Databases' suitability and case-study feasibility assessments followed three key steps: (I) characterization of data source systems and processes, using the EMA data quality checklist to evaluate foundational aspects and their maturity; (II) assessment of data quality metrics for each data source (data reliability), based on published research and open-access catalogues; and (III) fitness-for-use evaluation (data relevance), assessing database suitability for each case study based on question-specific determinants. Steps 1 and 2 were database-specific, while step 3 was both database- and case-specific, i.e. it could only be assessed in view of the specific research question to be addressed. From these steps, two tables containing qualitative information (I and III) and one with quantitative metrics (II) were created. The overall feasibility of the case studies using the candidate data sources was determined by critically analysing the collected information. Additional insights were gathered from DEAPs. All of the information was compiled into a report accompanying the generated tables, with our narrative assessment (appendix).

The overall feasibility assessment is summarized in Table 17. Both BIFAP and CPRD were deemed feasible data sources for studying dapagliflozin and MACE, with achievable sample sizes and reasonably up-to-date data. For BIFAP, the estimated sample size of ~13,341 participants is supported by its coverage of ~14 million inhabitants and frequent exposure to dapagliflozin. The database provides reliable hospitalization data and timely updates, but some limitations were identified: in-hospital cardiovascular events may not always be fully captured, mortality data is delayed by one year and lacks cause-of-death specificity, treatment discontinuation must be algorithmically inferred, and drug use is not linked to indication. These limitations may lead to some underestimation of outcomes and imprecision in time-sequencing, though major events are expected to be captured in primary care. CPRD also supports the target sample size and offers strong prescription data with good recency, but dispensing information is unavailable, discontinuation must be inferred from prescription duration, and diagnostic coding is incomplete in some emergency room settings. In addition, diabetes type is sometimes unspecified, though misclassification is expected to be minimal since insulin monotherapy can serve as a proxy for type 1 diabetes. Overall, while both sources present minor limitations, these are manageable within the study design and do not prevent the study from being feasible.

Table 17. Overall feasibility assessment summary for CS3 using BIFAP and CPRD

RWD source	Sample size estimation form the hypothetical trial protocol	Feasibility assessment *(yes/yes, with limitations/no)*	Rationale for the feasibility assessment	Limitations identified during the feasibility assessment and categorisation	Description of potential impact of the identified limitations on the study results
BIFAP	With an approximate estimated sample size of 13,341 (based on a 1:1 ratio of dapagliflozin and DPP-4i), and considering that BIFAP includes data from approximately 14 million inhabitants (up to 2018), the target sample size is anticipated to be reached. Furthermore, experimental exposure is expected to occur frequently (13.5 DHD in 2023) [1].	Yes	Elements with high criticality are available and fairly reliable. Data recency of 6 months before extraction, reasonably enough for the research question. The time elapsed from when a user requests the data to when they actually receive it is 1-4 months. Sample size is achievable.	Minor: Hospital information in BIFAP includes dates of admission and discharge, type of discharge, primary and secondary diagnoses at hospital discharge. So, an acute cardiovascular event will only be picked if it constituted one of the main reasons for admission. Minor: Mortality data is updated with a one-year delay relative to the present time. For research purposes only the year of death is available. Minor: In mortality, a non-random pattern of missingness (MNAR) was observed due to incomplete or inaccurate recording of the cause of death, with a tendency to preferentially register cardiovascular-related deaths. A non-random pattern of association between missingness and MACE was seen. GPs do not have a complete registry of deaths and, particularly, there is not an appropriate recording of the cause of death. Consequently, adjustments using statistical methods for MNAR should be considered in the TTE protocol. Minor: Discontinuation date is not available, but calculated by dispentation date+number of packages+posology if written by doctor, if not, calculated by algorithm.	As in-hospital cardiovascular events might not be fully captured, some underestimation of outcomes may exist. However, as these are usually severe and with chronic repercussions, we expect primary care setting will capture them even with some delay. As mortality data is delayed and only the year of death is available, this can impact precision and the time sequence of outcomes.

				Minor: Drug use is not linked to a specific indication. Minor: Smoking status may be biased, as the criterion is 'current use or use within one year prior to randomization; therefore, patients who smoked before this period would be classified as non-smokers.	
CPRD	With an approximate estimated sample size of 13,341 (based on a 1:1 ratio of dapagliflozin and DPP-4i), and considering that CPRD includes data from approximately 4.4 million inhabitants (as of 2014), the target sample size is anticipated to be reached. Furthermore, experimental exposure is expected to occur frequently.	Yes	Elements with high criticality are available and fairly reliable. Data recency of 3 months before extraction, reasonably enough for the research question. Sample size is achievable.	Minor: Dispensing is not available, only prescription. Minor: Treatment discontinuation is not readily available but inferred from prescription duration. Minor: Diagnostic codes are available for 86% subjects attending emergency room. Minor: Diabetes mellitus without type specification occurs frequently as well; usually insulin in monotherapy is used to assess T1D.	As this database only has prescription data, it is unknown if patients took the prescription or if they discontinued it. However, treatment duration is available, from which this data may be estimated. Diagnostic codes are reported to be available for 86% of subjects in the emergency room; however, the missing cases we expect to capture them from hospitalization records or primary care records, since the severity of this disease may justify an admission and/or the follow-up with the GP, or change of baseline treatment. As diabetes type is frequently not specified, insulin in monotherapy might be used as a proxy to detect T1D cases. Very low misclassification of indication is expected since insulin in monotherapy is not used for other indications rather than T1D.

Table 18. Metadata about data sources and software

	Data 1 (CPRD)	Data 2 (BIFAP)
Data Source(s):	CPRD GOLD and CPRD Aurum	BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria)
Study Period:	Expected: 2012–2025 (final dates TBD based on most recent data release)	Expected: 2012–2023 (final dates TBD based on availability)
Eligible Cohort Entry Period:	Expected: 2013–2024 (final dates TBD)	Expected: 2013–2022 (final dates TBD)
Data Version (or date of last update):	CPRD GOLD: November 2024 CPRD Aurum: September 2024	Latest available version: 2023 extract (annual update)
Data sampling/extraction criteria:	Patients ≥40 with T2D initiating dapagliflozin or DPP-4i with ASCVD or risk factors; ≥1 year lookback.	Patients ≥40 with T2D initiating dapagliflozin or DPP-4i with ASCVD or risk factors; ≥1 year lookback
Type(s) of data:	Primary care EHR, linked secondary care data (HES, ONS), prescriptions, lab values, demographics, outcomes	Primary care EHR, linked hospital discharge, mortality registry, pharmacy dispensation, vaccination data
Data linkage:	Yes: deterministic linkage with HES, ONS mortality, NCRAS, etc. (linkage via NHS number, DOB, postcode, sex)	Yes: deterministic linkage to mortality, hospitalisation, COVID vaccine/test data in some regions
Conversion to CDM*:	Yes: OMOP CDM, also CONCEPTION CDM	Yes: BIFAP CDM, also OMOP and CONCEPTION
Software for data management:	EMIS Web for Aurum, Vision for GOLD; analyses via CPRD secure TRE or local TRE	Comprehensive Study Management (GIE); AEMPS secure infrastructure
HMA data catalogue link	https://catalogues.ema.europa.eu/node/1026/administrative-details	https://catalogues.ema.europa.eu/node/955/administrative-details

7.8 Data Management

The study will be conducted in a distributed manner using the UMCU, ARS Toscana and VAC4EU tools, procedures, and pipeline. Figure 1 specifies the data sets (D) and transformation processes (T), programming follows this pipeline, with involvement of different types of experts.

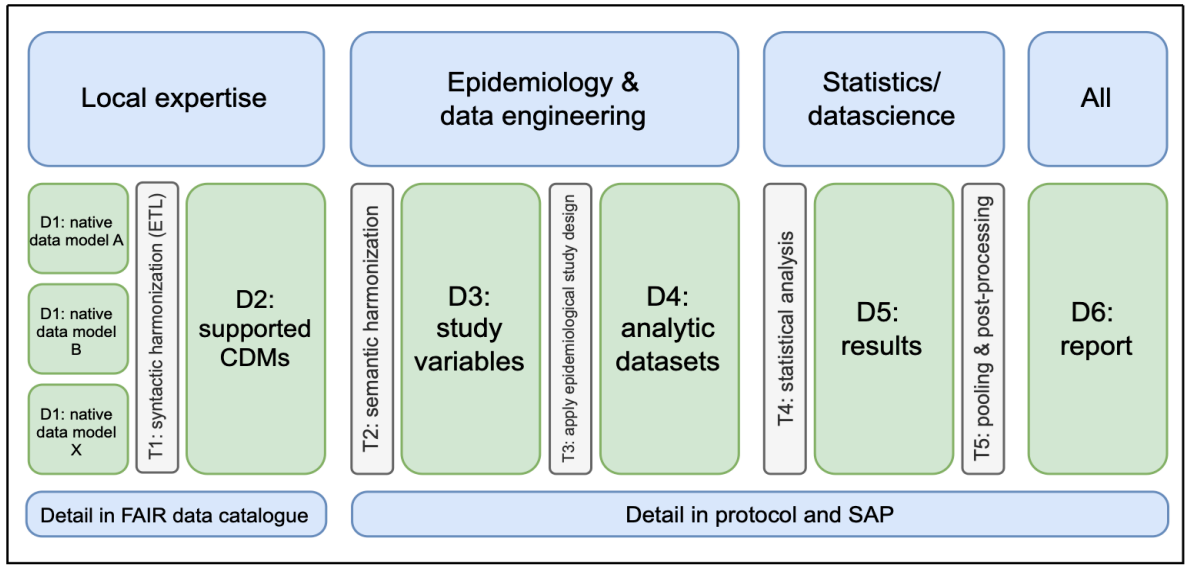


Figure 1. Data Management from the data transformation perspective

D1: Original data can be in any native format

The RWD-RWE pipeline used by VAC4EU starts with data banks that are controlled by the Data Expert and Access Partner (DEAP) and can be in any format. Data always stays local and never leaves the secure environments of the DEAPs. The ETL (extract, transform, load, see below for more details under 'T1') design is shared in a searchable FAIR VAC4EU catalogue. The VAC4EU FAIR Molgenis data catalogue is a meta-data management tool designed to contain searchable meta-data describing organisations that can provide access to specific data sources.

T1: Syntactic harmonisation (ETL)

T1: Syntactic harmonisation is conducted through an extraction, transformation, and loading (ETL) process of native data into the ConcePTION common data model (CDM) (see section 'D2: Common data model'). To harmonise the structure of the data sets stored and maintained by each data partner, a shared syntactic foundation is used. The ETL process has various structured steps as described by Thurin et al (2021)[9]

- DEAPs are asked to share the data dictionaries of their data banks (selected tables and variable names/structure)
- Metadata (descriptive data about the data sources and databanks) & data dictionaries, are uploaded in FAIR data catalogue (Molgenis).

D2: Common data model

For this project, the CDM (D2) is the ConcePTION common data model. The CDM version that is used is v2.2, which is available as an open-source CDM. In this CDM, data are represented in a common structure, but the values of the data remain in their original language (e.g. codes will have either ICD9/10/ICPC/SNOMED or MEDCODEID values).

T2: Semantic harmonisation

During the T2 step, many data transformations occur related to the completion of missing features in the data. Based on the relevant diagnostic medical codes and keywords, as well as other relevant concepts (e.g., medications), one or more phenotype algorithms are constructed (typically one sensitive, or broad, algorithm and one specific, or narrow, algorithm) to operationalise the identification and measurement of each event. In this step we conduct time anchoring (observation periods, look back periods), clean the data such as the dose of vaccines, sort on record level, aggregate across multiple records, and combine concepts for implantation of algorithms, and rule-based creation of study variables.

In this phase of the creation of study variables, semantic mapping is conducted. This semantic mapping across different vocabularies is conducted as part of the R-study script using different functionalities. To reconcile differences between different terminologies and native data availability, machine-readable code lists are used that comprise the terminologies that are used in the network (e.g. ICD-9, ICD10, SNOMED, ICPC and DEAP specific adaptations). This is combined with the BRIDGE metadata file that defines risk windows, look-back periods, and algorithms for each study variable (10).

D3: Study variables

D3 datasets are interim data sets with information on study variables for each study participant, the unit may be a person, a medicine, or episode of time. The design of these datasets is described in codebooks. Examples of D3 datasets are the outputs of the ConcePTION pregnancy algorithm (https://github.com/IMI-ConcePTION/ConceptionTools/wiki#conception-pregnancy-algorithm), and outputs of functions that define smoking. Multiple functions/packages exist within the VAC4EU, for different study variables.

T3: Application of epidemiological design

In the T3 step epidemiological designs are applied such as sampling, matching (on specific variables and/or propensity scores), and selection based on inclusion and exclusion criteria using the study variables in the D3 datasets. The designs will be implemented for the various study objectives using R-scripts, and these may use the existing functions (R-cran) or functions that have been developed in the VAC4EU community (e.g. matching).

D4: Analytical data set

D4 is an analytical dataset, and multiple D4 data sets may be produced based on the objectives of the study. The format is described initially in a code book for communication between programmers and statisticians.

T4: Statistical analysis

This step in the data transformation pipeline will produce statistical estimates such as descriptives (counts, percentages), distributions (mean, percentiles), rates (prevalence, incidence), regression coefficients, or other relevant estimates. This will be conducted using R.

D5: Results

D5 is the set of estimates, tables or aggregate data that is transferred from the DEAPs to the Digital Research Environment (DRE). The aggregated results produced by these scripts at the DEAP's site will be uploaded to the UMCU DRE for post-processing, pooling and visualisation (Figure 1). The DRE is a cloud-based, globally available research environment where data are stored and organised securely and where researchers can collaborate. The DRE is made available through UMCU. The DRE applies double authentication where researchers can collaborate using data that are stored and organised securely [ref]. UMCU is responsible for data processing and data security.

All researchers who need access to the DRE will be granted access to study-specific secure workspaces by UMCU. Access to the workspaces will be possible only after double authentication using an identification code and password together with the user's mobile phone for authentication.

Uploading files will be possible for all researchers with access to the workspace within the DRE. Downloading of files will be possible only after requesting and receiving permission from a workspace member with an "owner" role, who will be a UMCU team member.

T5: Post-processing/pooling

In this step, the result from different DEAPs is pooled and converted into tables and figures for reporting.

7.9 Quality Control

All key study documents such as the hypothetical trial protocol, target trial emulation protocol and study reports will undergo senior scientific and editorial review.

Data quality

For all data sources and for each data instance we will conduct *INSIGHT*[10] level 1-2 quality checks, detailed statistical analysis plans for the indicators are available on the public repositories:

- https://github.com/UMC-Utrecht-RWE/INSIGHT-Level1 Hoxhaj, V. (2023). UMC-Utrecht-RWE/INSIGHT-Level1: https://doi.org/10.5281/zenodo.10035167
- https://github.com/UMC-Utrecht-RWE/INSIGHT-Level2 Hoxhaj, V., & van den Bor, R. (2023). UMC-Utrecht-RWE/INSIGHT-Level2: https://doi.org/10.5281/zenodo.10035169

Briefly, level 1 verifies Data Completeness and level 2 Data Consistency.

Level 1 – Data Completeness

The purpose of the level 1 check is to verify the completeness of the ETL process and the data in the variables. Examples of tests are:

- Presence of variables in each of the CDM tables in D2
- Checks for misspellings and letter case in variable names in each of the CDM tables
- Verification of vocabularies
- Check date formats
- Check conventions of values
- Missing data analysis
- Frequency tables for categorical variables

Level 2 – Data Consistency

Real data is not random but follows certain logical constraints that reflect rules governing real-world situations. Examples of indicators generated by level 2 checks are:

- Event dates before date of birth
- Event dates after date of death
- Event dates out of observation periods
- Subjects having an observation but not present in the PERSONS table
- Observations associated with a visit id and occurred before/after the visit start/end date
- Subjects younger than 12 years old reported as parents
- Age at the observation period older than 115 y old Data

Code Quality

These coding practices define how the TARGET programming team collaborates to write clean, reliable, and reproducible code for the VAC4EU Real-World Evidence (RWE) Analytical Pipeline. They aim to ensure clarity, consistency, and maintainability across all case studies within the project.

Coding conventions

To ensure clarity, consistency, and maintainability across the project, the following conventions will be applied to all codebases within the project:

- Consistent style: Code follows a consistent and readable style (see the tidyverse style guide for R).
- Meaningful names: Use clear, descriptive names for variables, functions, and files to convey their purpose.
- Modular code: Break down code into small, reusable functions where possible.
- No hardcoded paths: Use configuration files or relative paths to ensure portability.

Following these conventions makes the code easier to understand, test, and reuse across case studies and teams.

Documenting Code

Code documentation is used to promote good coding practices and ensure our work is understandable, maintainable, and reproducible. To achieve this, we will:

- Use descriptive comments that explain the purpose and rationale behind code sections, focusing on why something is done, not just what.
- Clearly document function inputs, outputs, and side effects, using standardized formats (e.g., roxygen2 in R) where appropriate and supported.
- Write meaningful variable and function names to make the code as self-explanatory as possible.

Version Control

We use Git and GitHub to manage version control. These tools support good coding practices by enabling collaboration, tracking changes, accessing a project's history, and ensuring code quality through review and documentation.

A dedicated GitHub organisation has been created for the project (https://github.com/target-roc19). Each case study is managed in its own repository within this organisation. Repositories are structured consistently across case studies, to reinforce modularity. Access to repositories is controlled through teams.

During development, all repositories remain private to ensure confidentiality. Once the project is finalised, relevant repositories will be made public and assigned a digital object identifier (DOI) via Zenodo to support transparency, reproducibility, and reuse by the wider research community.

To maintain code quality and clarity, we follow the git and GitHub guidelines below.

- Always use pull requests (PRs): never push directly to the main branch.
- Open an issue before creating a new branch. Ideally, one PR resolves one issue to keep changes focused and reviewable.
- Every PR must be reviewed by at least one other person before merging.
- The PR author merges the PR after it has been reviewed and approved.
- Write clear, descriptive commit messages.
- Write informative PR descriptions, including:
    - A concise title
    - Links to related issues
    - A summary of the changes

Continuous Integration

Continuous Integration (CI) is set up to automatically check code quality and run tests whenever changes are pushed to the repository or submitted through a pull request (PR). The CI workflow ensures that the package adheres to predefined style guidelines and that all automated tests pass before changes are merged.

Coding Template

Every case study follows the general coding template used across all code in the TARGET project. The folder structure is organised as follows:

```
case-study-template
|____data
| |____D2_cdm
| |____D3_study_variables
| |____D4_analytic_datasets
| |____D5_results
| |____D6_report
|____docs
|____logs
|____run
|____tests
|____transformations
```

```
| |____T2_semantic_harmonization
| |____T3_study_design
| |____T4_statistical_analysis
| |____T5_processing_results
|____CHANGELOG.md
|____LICENSE
|____README.md
```

Project Data Structure and Storage

The data folder follows the Real-World Evidence pipeline structure. Data conforming to the common data model is stored in the D2_cdm folder.

Results from transformations T2, T3, T4, and T5 are saved in the respective folders:

- D3_study_variables
- D4_analytic_datasets
- D5_results
- D6_report

Each dataset is associated with a codebook, explained in more detail below.

All data remain securely stored on the Data Expert and Access Partners (DEAPs) servers and are never transferred externally. For testing purposes, dummy datasets are created. These fall into two categories:

- Unit test data: Small, predefined input and output pairs used to test individual transformation steps. These are stored in the tests folder, not in data, and can support automated testing.
- Pipeline test data: Larger, more complex dummy datasets used to test whether the full pipeline runs as expected. These may be included in the repository only if they remain below GitHub's 100 MiB file size limit and will otherwise be shared via SharePoint.

Logging System

When the pipeline is executed, log files are saved in the logs folder. These logs are especially helpful when running the code in the DEAPs environment, as they help trace and diagnose potential errors. We recommend using the logger R package to handle logging throughout the pipeline. A sample logging setup can be found in the logger.R script located at the root of the project directory.

Executing the Analytical Pipeline

The run folder contains scripts used to execute each transformation step in the pipeline.

- A central script, run_pipeline.R, orchestrates the full pipeline from start to finish.
- Subscripts (e.g., run_T2.R or similar) are available to run individual transformation steps separately.

Typically, the run_pipeline.R script is the main entry point used by a DEAP to execute the full pipeline. Before running it in the DEAP environment, the pipeline may need to be adapted to local settings. This can be done using a configuration file that defines variables required to tailor the pipeline to a specific DEAP. Please note that configuration files should not include sensitive information.

Such a file might include variables like:

- The name of the DEAP
- The path to the local data instance
- The path to any required external resources

Testing and Quality Assurance

The tests folder contains scripts to test the analytical pipeline. Tests will be used to ensure code behaves as expected and remains stable over time. By systematically checking inputs, outputs, and edge cases, tests help catch errors early and make future changes safer. We use the testthat R package to structure and run unit tests.

Continuous integration (CI) is used to automate testing. With CI, tests are automatically run each time code is pushed to the repository (e.g., via GitHub Actions). This helps identify issues immediately, ensures that new changes do not break existing functionality, and supports better collaboration by enforcing consistent code quality across contributors.

Modular Data Transformation Workflow

The transformations folder follows the Real-World Evidence pipeline structure. It contains the source code for all transformation steps, which is typically written in R. Each subfolder corresponds to a specific step in the pipeline (e.g., T2_semantic_harmonization, T3_study_design, T4_statistical_analysis, T5_processing_results) and includes the relevant scripts and helper functions for that step.

During the T2 step, a database is usually created (e.g., using DuckDB). This database can be queried using SQL, and it is recommended that all SQL queries be saved as clearly named, standalone SQL script files to ensure readability and reusability.

The purpose of the transformations folder is to structure and modularise the processing logic, making it easier to maintain, test, and reuse across different case studies. By organising code by transformation step, teams can work in parallel, increasing efficiency.

Changelog

A changelog will be kept for all notable changes in the project. Changelogs help track the evolution of the project over time, making it easier for collaborators to understand what has changed between versions. We follow the structure and best practices outlined in Keep a Changelog.

Codebooks

Before developing code, codebooks are created to describe each dataset (D) within the pipeline. A codebook is a comprehensive document that outlines the structure, contents, and metadata of a dataset. It serves as a detailed reference guide for anyone working with the data and plays a crucial role in guiding the development of the analytical pipeline by clearly defining both the inputs and expected outputs.

All codebooks are summarized in a central index file, which provides a high-level overview of the pipeline's structure. For each codebook, the index file includes:

- A brief description of its purpose,
- A list of the scripts used to generate the corresponding dataset,
- A description of the input datasets and input parameters required.

The datasets D2, D3, D4, and D5 are typically subdivided into multiple smaller transformation steps, each detailed within their respective codebooks. These smaller transformation steps ensure that each part of the pipeline is clearly scoped and well-documented.

In addition to supporting development, codebooks help ensure quality control by making transformation logic transparent and verifiable, and they enhance reproducibility by documenting exactly how data is structured and used throughout the analytical pipeline.

Deployment

The analytical pipeline is delivered to DEAPs as a GitHub release, tagged with a version number. Versioning follows the format: vYYYYMMDD.XX, where the date indicates the release date and XX denotes the sub-version or revision number.

Any deployment issues can be reported via the GitHub repository using the issues feature, where the programming team responsible for the R code will collaborate with the local DEAP to resolve them as needed.

Reproducibility

It is recommended to locally use the renv R package to maintain the R version and version of packages for reproducibility purposes.

At this time, however, using renv reliably across different systems and environments remains challenging. For this reason, we currently recommend its use only in local development setups.

We are actively monitoring developments in the R ecosystem related to cross-platform reproducibility. As soon as a more stable and portable solution becomes available, we will revisit this guidance and promote broader adoption.

Licensing

The code will be made available under an open source license.

README Guidelines

Each case study repository includes a README that covers the following points:

- Project Overview: brief summary of the study goals and key research questions.
- Background: context and rationale for the study.
- Repository Structure: Outline of main folders and their contents.
- Data Overview: Description of data sources, formats, and data privacy considerations.
- How to Run: Instructions for running the pipeline and key scripts, plus where outputs are saved.
- Testing: How to run tests to verify code functionality.
- Contributing: Guidelines for code contributions and issue tracking.
- License: Information about the code license.
- Contact: Who to reach out to for help or questions.

7.10 Study Precision

In this non-interventional study, no hypothesis test will be performed. The focus is on the precision of the estimated treatment effect. Assuming the study size in each RWD source will be similar to the sample size of the target trial, the level of precision that is achievable with a fixed sample size can be estimated using the expected width of the 95% confidence interval for the effect estimate.

Estimation

To estimate the expected 95% CI for a HR from a Cox proportional hazards model, the standard error (SE) of the log(HR) is derived from the total number of events.

Assumptions
- Equal allocation to treatment groups.
- Large-sample normal approximation for log(HR).
- Symmetric CI on the log scale.

The confidence interval (CI) width for the hazard ratio (HR) can be calculated using the following formula:

$CI\_width\_HR = exp(\hat{\beta} + 1.96 \times SE) - exp(\hat{\beta} - 1.96 \times SE)$

Where:
- $\hat{\beta}$ is the log hazard ratio (log(HR))
- SE is the standard error of the log(HR)
- 1.96 is the quantile 97.5% of a Normal(0, 1) distribution needed to construct a 95% confidence interval

Calculation of 95% CI

1. Assume equal allocation:
Number of events per group: d1 = d2 = d / 2

2. Calculate SE of log(HR):
SE[log(HR)] = sqrt(1/d1 + 1/d2) = sqrt(2/d)

3. Construct the 95% CI on log scale:
log(HR) ± 1.96 × SE[log(HR)]

4. Convert back to HR scale:
CI_HR = exp(log(HR) ± margin)

% Precision= (upper limit of the CI/ assumed HR) -1

Scenario 1
Calculation based on 1,189 events, HR=0.85

SE = sqrt(2 / 1189) = 0.04101

log(HR) = log(0.85) = -0.1625

Margin = 1.96 × SE = 1.96 × 0.04101 = 0.08038
Lower bound = -0.16252 - 0.08038 = -0.24290
Upper bound = -0.16252 + 0.08038 = -0.08214

Lower CI = exp(-0.24290) = 0.784
Upper CI = exp(-0.08214) = 0.921

% precision = (0.92/0.85)-1=8.23%

Scenario 2

Calculation under assumption that overall event rate for MACE is 10% lower than expected (1070), HR=0.85

SE = sqrt(2 / 1070) = 0.04323

log(HR) = log(0.85) = -0.16252

Margin = 1.96 × SE = 1.96 × 0.04323 = 0.08474
Lower bound = -0.16252 - 0.08474 = -0.24726
Upper bound = -0.16252 + 0.08474 = -0.07778

Lower CI = exp(-0.24726) = 0.78
Upper CI = exp(-0.07778) = 0.93

% precision=(0.93/0.85)-1=9.41%

Scenario 3

Calculation under assumption that overall event rate for MACE is 30% lower than expected (823), HR=0.85

log(HR) = -0.16252
Standard Error (SE) = 0.0493
Margin of Error = 0.09662
95% Confidence Interval (HR) = (0.772, 0.936)

% precision=(0.936/0.85)-1=10.11%

Table 19. % Precision under different events counts

Scenario	Number of Events	Hazard Ratio (HR)	log(HR)	Standard Error (SE)	Margin of Error	Lower CI (HR)	Upper CI (HR)	% precision
Scenario 1	1189	0.85	-0.1625	0.04101	0.08038	0.784	0.921	8.23
Scenario 2	1070	0.85	-0.1625	0.04323	0.08474	0.78	0.93	9.41
Scenario 3	823	0.85	-0.1625	0.0493	0.09662	0.72	0.936	10.11

8. Limitation of the methods

A key limitation of emulating a target trial using real-world data is that it may not be possible to fully reproduce all features of the target trial. These limitations are inherent to the target trial emulation approach.

Population, Eligibility, and Setting

In the target trial, tobacco use within the past year is an explicit risk factor for ASCVD. In the emulation, this factor is identified from routine healthcare data, where tobacco use is often under-recorded. This introduces a risk of misclassification, potentially excluding some individuals at high cardiovascular risk and attenuating comparability with the target trial population.

Eligibility criteria were operationalised using diagnostic and prescription codes with, while chronic conditions were measured over the entire observable history. These pragmatic definitions may lead to incomplete ascertainment of comorbidities and risk factors compared with trial screening procedures. As a result, some patients included in the emulation may not have met the target trial's eligibility thresholds. However, similar to the case of tobacco use, such misclassification is more likely to lead to the exclusion of potentially eligible patients with high ASCVD risk rather than the inclusion of patients who are not truly at high risk. The risk of this type of misclassification is lower than for tobacco use.

Treatment Conditions and Exposure

Background therapy is not restricted, reflecting real-world practice, but also introducing heterogeneity in treatment regimens. These factors represent departures from the controlled dosing and treatment monitoring procedures of a randomised trial.

Exposure is defined based on prescription records, which do not capture whether patients actually filled or took their medications. As a result, some individuals categorized as "exposed" may not have received the intended treatment. This limitation is acknowledged as a source of non-differential exposure misclassification, which would bias effect estimates toward the null. A sensitivity analysis will be implemented to assess the impact of varying levels of sensitivity and specificity in exposure classification, based on published estimates of prescription adherence.

Treatment Assignment and Follow-up

Randomisation cannot be emulated. Instead, inverse probability of treatment weighting is used to approximate exchangeability. While IPTW balances measured baseline covariates, it cannot account for unmeasured confounding. Moreover, trimming to address propensity score distribution non-overlap restricts the analysis to a subset of patients, meaning the estimated effects may not generalise to the entire eligible population.

In the emulation, follow-up begins at the date of first prescription rather than at randomisation. This aligns the start of follow-up with treatment initiation but differs from the conditions of the target trial, where treatment allocation is randomised, and follow-up begins at randomisation. In our

study, treatment allocation reflects real-world prescribing decisions based on patient characteristics rather than random assignment, which may introduce confounding. Furthermore, follow-up is determined by data availability and practice registration rather than scheduled study visits, which may result in imperfect measurement of loss to follow-up.

Outcomes and Intercurrent Events

Endpoints are identified using validated code lists in real-world data sources. Although these definitions are well established, they are not identical to adjudicated outcomes in clinical trials. Misclassification of myocardial infarction, stroke, or cause of death remains possible, which may bias effect estimates.

Strategies for handling intercurrent events (treatment discontinuation, switching, or addition of another glucose-lowering drug) are implemented using prescription records. While these reflect real-world effectiveness, the accuracy of operational definitions (e.g., gaps of >90 days to define discontinuation) is limited by prescribing and refill practices. For Estimand 2, in particular, misclassification of discontinuation or switching may bias estimates of while-on-treatment effects.

Loss to follow-up is defined using practice deregistration or database end, which are proxies for true loss. Although unique patient identifiers mitigate risks of missed de-registration, there remains potential for misclassification (e.g., patients who stop attending their practice but do not formally deregister).

Analysis Methods and Statistical Assumptions

Unlike the target trial, which relied on randomisation to achieve balance of baseline participant characteristics, the emulation employs IPTW in Cox and AFT models. This approach requires unverifiable assumptions: no unmeasured confounding, correct model specification, and positivity. Departures from these assumptions could bias results.

The Cox proportional hazards model assumes that the relative treatment effect is constant over time, which may not be appropriate. This assumption is explicitly tested using Schoenfeld residuals and log-log survival plots. In addition, a supplemental analysis is conducted using an accelerated failure time (AFT) model with a Weibull distribution to estimate restricted mean survival time (RMST) at fixed time points (3 and 5 years).

Some subgroups of patients may be very unlikely to receive one of the treatments (e.g., older patients with severe kidney disease may preferentially receive DPP-4i), leading to limited overlap in covariate distributions between groups. The propensity score distribution is examined for evidence of positivity violations by assessing the overlap of propensity scores between treatment groups. Trimming further narrows the study population, restricting generalisability.

Missing Data and Censoring

Unlike the trial, where exposure and covariates were actively collected, the emulation must rely on real-world data. Missing exposure data may occur through incomplete prescribing records, while covariates (e.g., lifestyle factors) are imputed under a missing-at-random assumption. Misclassification is also possible when absence of a diagnostic code is assumed to reflect absence of a condition. These differences from trial data collection procedures may affect validity.

A key limitation of this study is the reliance on assumptions regarding the censoring mechanism. In the primary Cox model, we assume that censoring is non-informative, meaning that the probability of being censored is independent of the outcome, conditional on treatment assignment, survival up to the time of censoring and indirectly on covariates included in the IPTW model . This assumption may not always hold in practice. For example, patients may leave the database or deregister from a practice due to worsening health, which could be directly related to their risk of experiencing the outcome. If this relationship is not fully captured by measured covariates, effect estimates could be biased. Even for the measured covariates, this indirect adjustment via IPTW would depend on how strongly censoring is associated with exposure. To address this, we perform sensitivity analyses using inverse probability of censoring weights (IPCW), which relax the primary assumption by directly adjusting informative censoring not only for baseline confounders but also for time-varying covariates that predict both censoring and the outcome. While IPCW provides a more flexible and potentially more realistic adjustment, it remains dependent on correct model specification and the availability of sufficient data to capture predictors of censoring. If important determinants of loss to follow-up are unmeasured or poorly recorded, residual informative censoring may persist. In addition, a tipping point sensitivity analysis is conducted to assess how strong the association between censoring and outcome would need to be to change the study's conclusions.

Data Source Heterogeneity

BIFAP and CPRD differ in population coverage, healthcare systems, coding practices, and linkage availability, which may introduce heterogeneity in effect estimates. Analyses will be performed separately within each data source using harmonized definitions under the Conception CDM framework. If pooled estimates are produced, heterogeneity will be assessed qualitatively.

9. Protection of human subjects

This is a non-interventional study using secondary data collection and does not pose any risks for individuals. Each data source research partner will apply for an independent ethics committee review according to local regulations. Data protection and privacy regulations will be observed in collecting, forwarding, processing, and storing data from study participants. Patient information This study involves data that exists in an anonymized structured format and contains no patient personal information. All parties will comply with all applicable laws, including laws regarding the implementation of organisational and technical measures to ensure the protection of patient personal data. Such measures will include omitting patient names or other directly identifiable data in any reports, publications, or other disclosures, except where required by applicable laws. Patient personal data will be stored at DAPs in encrypted electronic form and will be password protected to ensure that only authorised study staff have access. DAPs will

implement appropriate technical and organisational measures to ensure that personal data can be recovered in the event of a disaster. In the event of a potential personal data breach, DAPs shall be responsible for determining whether a personal data breach has in fact occurred and, if so, providing breach notifications as required by law.

Patient consent

As this study does not involve data subject to privacy laws according to applicable legal requirements, obtaining informed consent from individuals is not required.

10.Reporting of adverse events

For studies in which the research team uses only data from automated healthcare databases, according to the International Society for Pharmacoepidemiology Guidelines for GPP. “Aggregate analysis of database studies can identify an unexpected increase in risk associated with a particular exposure. Such studies may be reportable as study reports, but typically do not require reporting of individual cases. Moreover, access to automated databases does not confer a special obligation to assess and/or report any individual events contained in the databases. Formal studies conducted using these databases should adhere to these guidelines.” For non-interventional study designs that are based on secondary use of data, such as studies based on
medical chart reviews or electronic health records, systematic reviews, or meta-analyses, reporting of adverse events/adverse drug reactions is not required. Reports of adverse events/adverse drug reactions should only be summarized in the study report, where applicable. According to the EMA Guideline on GVP, Module VI – Management and Reporting of Adverse Reactions to Medicinal Products, “All adverse events/reactions collected as part of [non-interventional post authorization studies with a design based on secondary use of data], the submission of suspected adverse reactions in the form of [individual case safety reports] is not required. All adverse events/reactions collected for the study should be recorded and summarized in the interim safety analysis and in the final study report.” Module VIII – Post-Authorization Safety Studies echoes this approach. Legislation in the EU further states that for certain study designs such as retrospective cohort studies, particularly those involving electronic health records, it may not be feasible to make a causality assessment at the individual case level.

11. References

1. Wiviott SD, Raz I, Bonaca MP, et al. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med.* 2019;380(4):347–357. doi:10.1056/NEJMoa1812389

2. Green JB, Bethel MA, Armstrong PW, et al. Effect of Sitagliptin on Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med.* 2015;373(3):232–242. doi:10.1056/NEJMoa1501352

3. Scirica BM, Bhatt DL, Braunwald E, et al. Saxagliptin and Cardiovascular Outcomes in Patients with Type 2 Diabetes Mellitus. *N Engl J Med.* 2013;369(14):1317–1326. doi:10.1056/NEJMoa1307684

4.European Medicines Agency. EMA Regulatory Science to 2025 – Strategic Reflection. 2020. [Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf]

5. Cosentino F, Grant PJ, Aboyans V, et al. 2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. *Eur Heart J.* 2020;41(2):255–323. doi:10.1093/eurheartj/ehz486

6. American Diabetes Association. 10. Cardiovascular Disease and Risk Management: Standards of Medical Care in Diabetes—2024. *Diabetes Care.* 2024;47(Suppl 1):S162–S185. doi:10.2337/dc24-S010

7. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. Stat Med. 2018 Jun 30;37(14):2252-2266. doi: 10.1002/sim.7654.

8. Lemstra M, Nwankwo CK, Bird Y, Moraros J. Primary nonadherence to chronic disease medications: a meta-analysis. *Patient Prefer Adherence*. 2018;12:721-731. https://doi.org/10.2147/PPA.S161151

9. Thurin NH, Pajouheshnia R, Roberto G, Dodd C, Hyeraci G, Bartolini C, et al. From Inception to ConcePTION: Genesis of a Network to Support Better Monitoring and Communication of Medication Safety During Pregnancy and Breastfeeding. Clin Pharmacol Ther. 2022 Jan;111(1):321–31.

10. Hoxhaj V, Andaur Navarro CL, Riera-Arnau J, Elbers RJ, Alsina E, Dodd C, et al. INSIGHT: A Tool for Fit-for-Purpose Evaluation and Quality Assessment of Observational Data Sources for Real World Evidence on Medicine and Vaccine Safety [Internet]. Epidemiology; 2023 [cited 2024 Oct 24]. Available from: http://medrxiv.org/lookup/doi/10.1101/2023.10.30.23297753

11. Royo AC, Elbers Jhj R, Weibel D, et al. Real-World Evidence BRIDGE: A Tool to Connect Protocol With Code Programming. Pharmacoepidemiol Drug Saf. 2024;33(12):e70062. doi:10.1002/pds.70062