Study Report

1. Title Page

Title	TARGET-EU: Dapagliflozin and major adverse cardiovascular events in type 2 diabetes
Research question & 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.
Contributors	**Primary investigator contact information:** Dr. Daniala Weir (Utrecht University) d.l.weir@uu.nl **Contributor names and affiliations:** Prof. Olaf Klungel (Utrecht University) Prof. Ian Douglas (London School of Hygiene and Tropical Medicine) Dr. Mar Martin Perez (Agencia Española de Medicamentos y Productos Sanitarios) Dr. Belén Castillo Cano (Agencia Española de Medicamentos y Productos Sanitarios) Dr. Elisa Martin Merino (Agencia Española de Medicamentos y Productos Sanitarios) Dr. Patrick Souverein (Utrecht University) Dr. Tomas Lasys (Utrecht University) Dr. Rosa Ginni (ARS Tuscany) Davide Messina (ARS Tuscany) Olga Paoletti (ARS Tuscany) Claudia Bartolini (ARS Tuscany)
Study registration	**Site:** https://catalogues.ema.europa.eu/node/4883/administrative-details **Identifier**: EUPAS1000000979 *This study is part of the TARGET EU project with registration details:* **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
Disclaimer and acknowledgement	This document has been produced by making use of the results under the specific contract no.04 FWC EMA/2020/46/TDA/ L5.06 between EU PE&PV Research Network (coordinated by Utrecht University), and the European Medicines Agency. This document expresses the opinion of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties. This study has been registered in the HMA-EMA Catalogue of Real-World Data under the EU PAS numbers: EUPAS1000000539, EUPAS EUPAS1000000979.
Abstract	**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 first specified the estimand using the Estimand framework, then developed a target trial to align with this estimand, and emulated the target trial in a non-interventional study using 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 2013 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. **Results:** The final cohorts comprised 34,110 patients in CPRD (3,850 dapagliflozin, 30,260 DPP4-i) and 205,993 in BIFAP (71,159 dapagliflozin, 134,834 DPP4-i). After IPTW, adequate covariate balance was achieved. Crude incidence rates of MACE were higher for dapagliflozin than DPP4-i in CPRD (237.2 vs 169.8 per 1,000 person-years) but lower for dapagliflozin in BIFAP (57.1 vs 110.3 per 1,000 person-years). Under the primary estimand, dapagliflozin was associated with a higher hazard of MACE than DPP4-i in CPRD (adjusted HR 1.30, 95% CI 1.20–1.40) but a lower hazard in BIFAP (adjusted HR 0.62, 95% CI 0.61–0.64). This directional discordance between databases persisted across the supplementary estimand, RMST analysis, multiple imputation, and probabilistic bias analysis for exposure misclassification. **Conclusion:** The association between dapagliflozin and MACE relative to DPP-4 inhibitors differed in direction between CPRD and BIFAP and was not explained by measured confounding. This cross-database heterogeneity most likely reflects setting-specific differences in patient populations and outcome measurement (particularly the composite strategy used to capture all-cause mortality, which assumed predominantly cardiovascular deaths but, given the pandemic-overlapping study period, likely captured substantial COVID-19-related mortality) rather than a consistent causal effect of dapagliflozin on cardiovascular risk relative to DPP-4 inhibitors.

2. Table of Contents

1. Title Page 1
2. Table of Contents 4
3. List of Tables and Figures 6
4. Abbreviations 9
5. Protocol Amendments 11
6. Background 16
7. Research questions and objectives 17
    7.1. Primary Estimand 1 17
    7.2. Supplementary Estimand 2 18
    7.3. Supplementary Estimand 3 20
8. Research methods 20
    8.1. Study design 20
    8.2. Study design diagram 21
    8.3. Setting 22
        8.3.1. Definition of time 0 (and other primary time anchors) for entry to the study population 22
        8.3.2. Study inclusion criteria 23
        8.3.3. Study exclusion criteria 25
    8.4. Variables 27
        8.4.1. Exposure(s) of interest 27
        8.4.2. Outcome(s) of interest 29
        8.4.3. Follow-up 31
        8.4.4. Covariates (confounding variables and effect modifiers, e.g. risk factors, comorbidities, comedications) 32
    8.5. Core Emulation Table – Design Summary 35
    8.6. Data analysis 38

8.6.1. Analysis Plan 38
8.6.2. Primary Estimand (1) Analysis 39
8.6.3. Supplemental Estimand (2) Analysis 43
8.6.4. Supplemental Estimand (3) Analysis 44
8.6.5. Sensitivity Analyses 44
8.6.6. Other Supplemental Analyses 49
8.6.7. Core Emulation Tables – Estimation Summary 50
8.7. Data sources 52
8.7.1. Data sources and Quality 52
8.8. Data Management 58
8.9. Quality Control 61
9. Deviations from Protocol 68
10. Results 69
10.1. Study population 69
10.2. Patient characteristics 69
10.3. Outcome Descriptive Summary Statistics 83
10.4. Primary Estimand 98
10.4.1. Primary analyses 98
10.4.2. Sensitivity analyses 99
10.5. Supplemental Estimand(s) 100
11. Additional Analyses 101
12. Discussion 101
12.1. Key Results 101
12.2. Interpretation of results 102
12.3. Limitations 105
13. Conclusion 106

14. References 107
15. Appendices 109
15.1. Supplemental Tables and Figures 109
15.2. Precision-Based Sample Size Rationale 135
15.3. Study Meta data (code lists, code books) 137

3. List of Tables and Figures

List of Tables

Table 1. Protocol Amendments 11
Table 2. Estimand 1 17
Table 3. Estimand 2 18
Table 4. Operational Definition of Time 0 (index date) and other primary time anchors 22
Table 5. Operational Definitions of Inclusion Criteria 23
Table 6. Operational Definitions of Exclusion Criteria 26
Table 7. Operational Definitions of Exposure 28
Table 8. Operational Definitions of Outcome 29
Table 9. Operational Definitions of Follow Up 31
Table 10. Operational Definitions of Covariates 32
Table 11. Comparison of Target Trial and Proposed Target Trial Emulation Design Elements 35
Table 12. Sensitivity analyses -Inverse Probability of Censoring Weighting (IPCW) 44
Table 13. Sensitivity analyses -Tipping Point Analysis 46
Table 14. Sensitivity analyses -Non-Differential Exposure Misclassification 47
Table 15. Core Emulation Table Estimand 1: Estimation Summary 50
Table 16. Core Emulation Table Estimand 3: Estimation Summary 52
Table 17. Overall feasibility assessment summary for CS3 using BIFAP and CPRD 54
Table 18. Metadata about data sources and software 56
Table 19. Baseline characteristics of the study population overall and by treatment group (dapagliflozin vs DPP-4 inhibitors), before and after IPTW, CPRD 73
Table 20. Baseline characteristics of the study population overall and by treatment group (dapagliflozin vs DPP-4 inhibitors), before and after IPTW, BIFAP 78
Table 21. Intercurrent events, and reasons for end of follow-up by treatment group (dapagliflozin vs DPP-4 inhibitors) for MACE, per estimand, CPRD 88

Table 22. Intercurrent events, and reasons for end of follow-up by treatment group (dapagliflozin vs DPP-4 inhibitors) for MACE, per estimand, BIFAP90
Table 23. Crude incidence rates (per estimand) and restricted mean survival time for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors), CPRD....92
Table 24. Crude incidence rates (per estimand) and restricted mean survival time for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors), BIFAP....93
Table 25. Crude and adjusted hazard ratios (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 1), CPRD....98
Table 26. Crude and adjusted hazard ratios (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 1), BIFAP....98
Table 27. Adjusted hazard ratios (95% CI) for MACE comparing dapagliflozin with DPP-4 inhibitors across sensitivity analyses related to missing data, CPRD....99
Table 28. Adjusted hazard ratios (95% CI) for MACE comparing dapagliflozin with DPP-4 inhibitors across sensitivity analyses related to other modelling assumptions, CPRD, BIFAP....99
Table 29. Crude and adjusted hazard ratios (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 2), CPRD....100
Table 30. Crude and adjusted hazard ratios (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 2), BIFAP....101
Table 31. Crude and adjusted differences in restricted mean survival time (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 3), CPRD....101
Table 32. Crude and adjusted differences in restricted mean survival time (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 3), BIFAP....101

List of Figures

Figure 1. Study design diagram....21
Figure 2. Data Management from the data transformation perspective....58
Figure 3. Flowchart of patient exclusions, CPRD....71
Figure 4. Flowchart of patient exclusions, BIFAP....72
Figure 5. Crude cumulative incidence curves for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors) under Estimand 1, CPRD....94
Figure 6. Crude cumulative incidence curves for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors) under Estimand 1, BIFAP....95
Figure 7. Adjusted cumulative incidence curves for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors) under Estimand 1, CPRD....96
Figure 8. Adjusted cumulative incidence curves for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors) under Estimand 1, BIFAP....97

List of Supplemental Tables

Supplemental Table 1. Attrition Table, CPRD....109
Supplemental Table 2. Attrition Table, BIFAP....109
Supplemental Table 3. Baseline patient characteristics of the study population before imputation, CPRD....110
Supplemental Table 4. Patient Characteristics of the study population before imputation, BIFAP....114

Supplemental Table 5. Description of the propensity score model parameters used to compute weights in the final model, CPRD 117
Supplemental Table 6. Description of the propensity score model parameters used to compute weights in the final model, BIFAP 120
Supplemental Table 7. IPT weights distribution, by treatment arm, before and after truncation at 1 at 99% for CPRD..... 123
Supplemental Table 8. IPT weights distribution, by treatment arm, before and after truncation at 1 at 99% for BIFAP 124
Supplemental Table 9. % Precision under different events counts 137

List of Supplemental Figures

Supplemental Figure 1. Distribution and diagnostics of propensity scores, CPRD 125
Supplemental Figure 2. Distribution and diagnostics of propensity scores, BIFAP 126
Supplemental Figure 3. Schoenfeld Residuals Plot: Proportional Hazards Assumption for Dapagliflozin vs. DPP-4 Inhibitors (Outcome: MACE), CPRD127
Supplemental Figure 4. Schoenfeld Residuals Plot: Proportional Hazards Assumption for Dapagliflozin vs. DPP-4 Inhibitors (Outcome: MACE), BIFAP 128
Supplemental Figure 5. Log-Log Survival Plot for Dapagliflozin vs. DPP-4 Inhibitors; Assessment of the Proportional Hazards Assumption for MACE, CPRD..... 129
Supplemental Figure 6. Log-Log Survival Plot for Dapagliflozin vs. DPP-4 Inhibitors; Assessment of the Proportional Hazards Assumption for MACE, BIFAP 130

4. Abbreviations

Abbreviation	Definition
AFT	Accelerated failure time
ASDs	Standardised mean differences
ATE	Average Treatment Effect
ATT	Average treatment effect in the treated
ATU	Average treatment effect in the untreated
BOM	Beoordeling van Oncologische Middelen
CapOx	Capecitabine-oxaliplatin
CBS	Centraal Bureau voor de Statistiek
CDM	Common data model
CI	Confidence interval; Continuous Integration
CRC	Colorectal cancer
DEAP	Data Expert and Access Partner
DOI	Digital object identifier
DRE	Digital Research Environment
DWH	Datawarehouse
HER	Electronic health records
ESMO	European Society of Medical Oncology
ETL	Extract, transform, load

HIPEC	Hyperthermic intraperitoneal chemotherapy
HR	Hazard ratio
IPTW	Inverse probability of treated weighted
IQR	Interquartile ranges
PBA	Probabilistic Bias Analysis
PFS	Progression-free survival
PRs	Pull requests
MAR	Missing at random
MICE	Multiple imputation with chained equations
mCRC	Metastatic colorectal cancer
NCR	Netherlands Cancer Registry
NIS	Non-interventional study
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OS	Overall survival
RCTs	Randomised controlled trials
RMST	Restricted mean survival time
RWE	Real World Evidence
RWD	Real-world data
SACT	Systemic anticancer therapy

SMR	Standardised mortality ratio
VEGF	Vascular Endothelial Growth Factor

5. Protocol Amendments

Table 1. Protocol Amendments

Version date	Version number	Section of protocol	Amendment or update	Reason
27 February 2026	1.0			
4 June 2026	2.0			
		Title	Title shortened by removing "Comparative effectiveness and safety studies using the target trial emulation and estimand frameworks"	Harmonisation using the same titling convention as for all other case studies within the TARGET EU project.
		7.7.3 Section 7.5 and Table 11	Exclusion criteria added to restrict the study population to patients with hospital linkage. The change has also been reflected in the core emulation table of the study design.	Under specification: This was not explicitly stated and is important to clarify for CPRD studies.
		7.2, 7.4.1	Added in footnote to study diagram: "*Patient episode eligibility: a patient may be queried across multiple treatment episodes. If a patient does not meet all inclusions criteria during an earlier episode but does so in a later episode (after	Under specification: It is important to clarify that a patient may have multiple treatment episodes, and that only the first eligible episode will be selected for analysis.

Version date	Version number	Section of protocol	Amendment or update	Reason
			completing a washout), the first eligible episode defines T_0. The patient appears only once in the attrition table and patient flowchart, regardless of how many ineligible episodes preceded cohort entry." Also added to algorithm to define exposure: A patient may be queried across multiple treatment episodes. If a patient does not meet all inclusions criteria during an earlier episode but does so in a later episode (after completing a washout), the first eligible episode defines T_0	
		Table 5 footnote	The definition of *active prescription* has been changed from "Prescription within 30 days prior to index with coverage overlapping the index date" to: "any prescription of an antihypertensive medication in the 120 days prior to index"	The former criterion of a treatment duration lasting 30 days prior to the index date was revised for efficiency. Because treatment duration is not consistently recorded, the maximum possible duration derived from days' supply is used instead.
		Table 6	Change introduced to reflect that ACS and T1D diagnosis could be in any diagnostic position rather than primary or secondary diagnosis	Diagnosis position in CPRD HES does not correspond to primary, secondary, or other diagnosis types; therefore, restricting by position could miss relevant diagnoses.

Version date	Version number	Section of protocol	Amendment or update	Reason
		7.4.1	Replaced “standard dosing consistent with routine clinical practice” with “fixed duration of 30 days if dosing information incomplete	A more pragmatic approach that is easier to implement from an analytical perspective.
		7.4.1 and table 9	Added details on how the intercurrent events of *treatment switch* and *addition of another antihyperglycemic agent* will be identified using RWD: “An add-on was defined as initiation of a new glucose-lowering therapy while the index treatment episode remained active — that is, while a subsequent dispensing of the index treatment occurred within the 90-day grace period following the end of the previous prescription's days supply. A switch was defined as initiation of a new therapy after the index treatment episode had ended, indicated by the absence of a subsequent index treatment dispensing within that grace period. End of treatment was therefore incorporated into the add-on versus switch classification rather than treated as a competing censoring event, to avoid masking treatment switching. When the end of a treatment episode (defined as the last prescription's days supply plus	The previous definition was underspecified and did not provide sufficient detail to distinguish between the addition of an antihyperglycemic agent and a treatment switch. This distinction has now been clarified. The same clarification was also incorporated into Table 9.

Version date	Version number	Section of protocol	Amendment or update	Reason
			the 90-day grace period) coincided with a new dispensing of the same drug on the same day, the episode was considered continuous. When it coincided with a dispensing of a different drug on the same day, this was classified as discontinuation and switching, as the grace period for the index treatment had already elapsed by that date.”	
		Table 9	Split “end of observation data” into “Date of final data capture in data source” and “End date of individual observation in the data source (e.g., migration or de-registration from the practice)	The definition of end of observation was not sufficiently specific, as it could occur through two distinct mechanisms.
		7.6.2	Changed SMD to “Evaluate absolute standardized differences (ASDs): ASDs < 0.1 will be considered acceptable. Categorical variables will be treated as multiple binary indicators, and one ASD will be reported per level of the variable across all case studies.	The original description did not specify that absolute differences would be used and lacked detail on how categorical variables with multiple levels would be handled.
		Table 10	Frailty removed from list of covariates. This was also explained in core design and estimation tables.	Not possible to measure in BIFAP.
		7.7.6	Added “For all estimands, subsequent IEs will be reported only if the preceding IE does not result in censoring. “	The text was revised to clarify that patients can have multiple intercurrent events, while providing greater specificity regarding the

Version date	Version number	Section of protocol	Amendment or update	Reason
				events that will be included in the descriptive supplementary analyses.
		Annex	Inclusion of the study code list as annex	Inclusion of the study code list for programming activities to identify eligibility criteria, exposures, outcome, covariates, etc., which was missing.

6. 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]

7. 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.

7.1. *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). We restrict to patient population for which we have hospital linkage	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	

Attribute	Target Trial	Target Trial Emulation	Comment
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.

7.2. *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		

Attribute	Target Trial	Target Trial Emulation	Comment
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 a treatment switch from the addition of another antihyperglycaemic agent to the index therapy, which is defined as the initiation of another agent during continuous treatment with the index therapy. All cause death measured using death registry data	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.

7.3. *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?

8. Research methods

8.1. *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.

8.2. *Study design diagram*

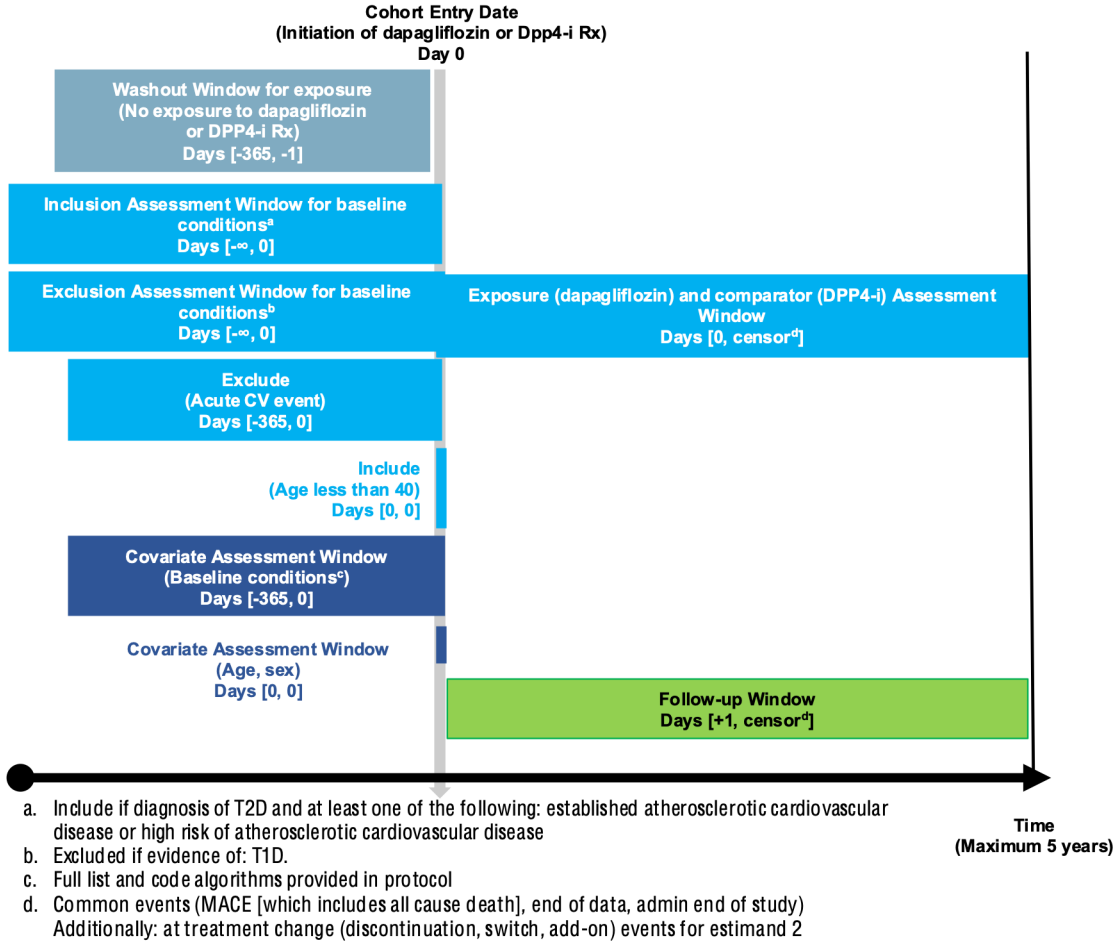


Figure 1. Study design diagram

*Patient episode eligibility: a patient may be queried across multiple treatment episodes. If a patient does not meet all inclusions criteria during an earlier episode but does so in a later episode (after completing a washout), the first eligible episode defines T_0. The patient appears only once in the attrition table and patient flowchart, regardless of how many ineligible episodes preceded cohort entry.

8.3. *Setting*

This study is conducted using routinely collected electronic health records from 2012 to 2024, a period starting after the introduction of dapagliflozin into 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 el Ámbito Público [BIFAP]), providing population-based and representative coverage of real-world clinical care.

8.3.1. *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 after the fulfilment of eligibility. This is when patients enter the study population and mimics the timing of initiation of therapy at randomization in the target trial.

A patient may be queried across multiple treatment episodes. If a patient does not meet all inclusions criteria during an earlier episode but does so in a later episode (after completing a washout), the first eligible episode defines T_0

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

8.3.2. *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	Appendix found in zenodo	Appendix found in zendodo
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	[-120, 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	[-120, 0]	OP	ATC	n/a	All study populations		

Criterion	Details	Assessment window	Care Settings[1]	Code Type[2]	Diagnosis position[3]	Applied to study populations:	Measurement characteristics/validation	Source for algorithm
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 any prescription of an antihypertensive medication in the 120 days prior to index

8.3.3. *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

Do not have linkage to hospitalization data

Rationale: Our base population in both data sources consists of patients registered in primary care. Outcomes can be identified using both hospital and primary care data when the corresponding diagnosis is available in both sources, but this is only possible for patients for whom we have access to linked hospital and primary care data. For patients with only primary care data available, we cannot determine with certainty whether the absence of an outcome reflects a true negative or simply incomplete capture of events occurring in hospital care.

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 Type[2]	Diagnosis position[3]	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	Appendix found in zendodo	Appendix found in zendodo
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	Any	All study populations		
Type 1 Diabetes	Diagnosis of type 1 diabetes at any point before index	[-∞, 0]	IP,OP	ICD-10, 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)

8.4. *Variables*

8.4.1. *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 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 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.

Follow-up for dapagliflozin and Dpp-4i users is capped at five years.

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, a fixed duration of 30 days will be used. 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.

An add-on was defined as initiation of a new glucose-lowering therapy while the index treatment episode remained active — that is, while a subsequent dispensing of the index treatment occurred within the 90-day grace period following the end of the previous prescription's days supply. A switch was defined as initiation of a new therapy after the index treatment episode had ended, indicated by the absence of a subsequent index treatment dispensing within that grace period. If the index treatment episode ended without initiation of any new glucose-lowering therapy, this was classified as discontinuation without switching. End of treatment was therefore incorporated into the add-on versus switch classification rather than treated as a competing censoring event, to avoid masking treatment switching. When the end of a treatment episode (defined as the last prescription's days supply plus the 90-day grace period) coincided with a new dispensing of the same drug on the same day, the episode was considered continuous. When it coincided with a dispensing of a different drug on the same day, this was classified as discontinuation and switching, as the grace period for the index treatment had already elapsed by that date. When no subsequent dispensing of any glucose-lowering therapy occurred, this was classified as discontinuation without switching.

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	Appendix found in zenodo	Appendix found in zenodo
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)

8.4.2. *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	Appendix found in zenodo	Appendix found in zenodo
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	Yes (as part of composite MACE)	Time-to-event	[-365;0]	IP, OP	ICD-10, SNOMED	Primary or secondary	Exposure: Dapagliflozin,		

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
	hospital discharge diagnoses. Hospitalization can also be recorded in Primary care records.							Comparator: DPP4-I		
All-cause mortality	Death from any cause, as recorded in mortality registry	Yes (as part of composite MACE)	Time-to-event		Mortality registry or, if that is not available,from Primary care records: the date on which the health insurance card was deregistered due to death	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)

8.4.3. *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
Date of final data capture in data source	Yes	Censor at database end; administrative censoring
End date of individual observation in the data source (e.g., migration or leaving the practice)	Yes	Censor at date of last healthcare contact, known deregistration; Non-administrative censoring
Day X following index date *(specify day)*	Yes	For estimand 1 & 2 Day 1825 (5 years after index date); administrative censoring. For estimand 3 1095 (3 years).
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. Censoring at treatment end occurs when a patient had no subsequent prescription within 90 days after the calculated end of the previous prescription's days' supply, with follow-up ending 90 days after the final prescription coverage period.
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. Censoring occurs at the prescription of the second drug. Additional follow-up days are then required to classify the event as add-on or switch.

8.4.4. *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							Appendix found in zenodo	Appendix found in zenodo
Age at index		Continuous	[0,0]			N/A		
Gender		Binary	[0,0]			N/A		
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		

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type[2]	Diagnosis Position[3]	Measurement characteristics/ validation	Source for algorithm
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			

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type[2]	Diagnosis Position[3]	Measurement characteristics/ validation	Source for algorithm
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			

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type[2]	Diagnosis Position[3]	Measurement characteristics/ validation	Source for algorithm
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)

8.5. *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)	Inclusion: Same: - Diagnosis codes for type 2 diabetes - Recorded ASCVD or ≥1 CV risk factors in baseline data	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

Attribute	Target Trial	Target Trial Emulation	Comment
	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 - No hospital linkage available (CPRD)	- 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 - 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.	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.
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	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.

Attribute	Target Trial	Target Trial Emulation	Comment
		antihyperglycemic therapy the patient may already be prescribed.	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. Frailty was initially identified as an important confounder but could not be reliably measured in both RWD sources.	Randomisation cannot be directly emulated. IPTW will be used in the statistical analysis to balance confounders in absence of randomisation. Inability to measure frailty can be considered a deviation from emulating random treatment assignment as it may induce residual bias.
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	Patients with known deregistration date, practice	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

Attribute	Target Trial	Target Trial Emulation	Comment
	condition and vital status remains unknown despite multiple attempts to contact them.	withdrawal or database end. This is directly measured in RWD source.	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

8.6. *Data analysis*

8.6.1. *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, and will include 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.

8.6.2. *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 in the analysis 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 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 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).

- **IPTW distribution check:** Evaluate the distribution of IPTW for each treatment arm (graphically), 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.

8.6.3. *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.

8.6.4. *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.

8.6.5. *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, the covariates entering the model will be updated, and a binary indicator will be assessed for whether the patient was censored during that interval. Patients who are not censored continue to contribute follow-up time to the risk set until they experience the event, are censored, or reach the end of follow-up. A separate conditional probability of remaining uncensored (non-administratively censored) 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 missingness not at random (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 directly 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/MAR), 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.

8.6.6. *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.

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

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.

For all estimands, subsequent IEs will be reported only if the preceding IE does not result in censoring.

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

8.6.7. *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 PS-based 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

Attribute	Target Trial	Target Trial Emulation	Comment
			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 be conducted 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 are 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).	Potential for exposure mismeasurement only present in emulation since exposure based on prescription records and assume adherence to prescribed treatment.

Attribute	Target Trial	Target Trial Emulation	Comment
		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	

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 15
Missing Data Assumptions and Methods to Handle	Same as Table 15		
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		

8.7. *Data sources*

8.7.1. *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 linkage to hospital discharge registries across all autonomous communities, while linkage to the cause of death and confirmed death date (national mortality registry) is available only in several regions. 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 (mean ~10 years for BIFAP [https://bifap.aemps.es/statistics?lang=en] and ~6–13 years for CPRD [https://www.cprd.com/data]) 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) as well as for hospital discharge data. Cause of death data are available through mortality registry for a large proportion of patients. 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.

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, lab values) may have inconsistent completeness or be missing entirely for subpopulations. . 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: https://bifap.aemps.es/data-governance?lang=en.

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: cause-of-death data is delayed by one year and lacked for 4 out of 12 regions, 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,	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	Minor: Cause of death data is available for 8 out of 12 regions and 8 years out of 12 years of the study period that are linked to the national Mortality registry. For the remaining regions/periods, the administrative death is used instead, i.e. the date on which the health insurance card was deregistered due to death is assumed as the date of death. Minor: In mortality data, we assume (rather than observe) a non-random pattern of missingness in cause-of-death recording consistent with a missing-not-at-random	As cause-of-death data is not available for the last 3 years of the study, i.e. 2021-2023

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
	experimental exposure is expected to occur frequently (13.5 DHD in 2023) [1].		actually receive it is 1-4 months. Sample size is achievable.	(MNAR) mechanism — that is, the completeness or accuracy of cause-of-death recording is hypothesized to depend on the true (unobserved) cause of death itself, rather than only on measured common causes, with a presumed tendency to preferentially register cardiovascular-related deaths. A non-random pattern of association between missingness and MACE was seen. Cause of death is informed by the national Mortality Registry — a complete registry — for the majority of the study period (2013–2020) and regions; outside this coverage, however, recording of the cause of death may be less complete. Consequently, adjustments using statistical methods for MNAR should be considered in the TTE protocol for the periods and regions not covered by the Mortality Registry linkage. Minor: Discontinuation date is not available, but calculated by dispentation date+number of packages+posology if written by doctor, if not, calculated by algorithm. Minor: Drug use is not linked to a specific indication. Minor: Smoking status may be biased, as 'smoker' or 'former smoker' diagnostic records were used for categorization within one year prior to randomization; therefore, patients without those diagnostic records who smoked before this period would be classified as non-smokers.	

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
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 el Ámbito Público)
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: 2024 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 cause of death registry in some regions, hospitalisation
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

8.8. *Data Management*

The study will be conducted in a distributed manner using the UMCU, ARS Toscana and VAC4EU tools, procedures, and pipeline. Figure 2 specifies the data sets (D) and transformation processes (T), programming follows this pipeline, with involvement of different types of experts.

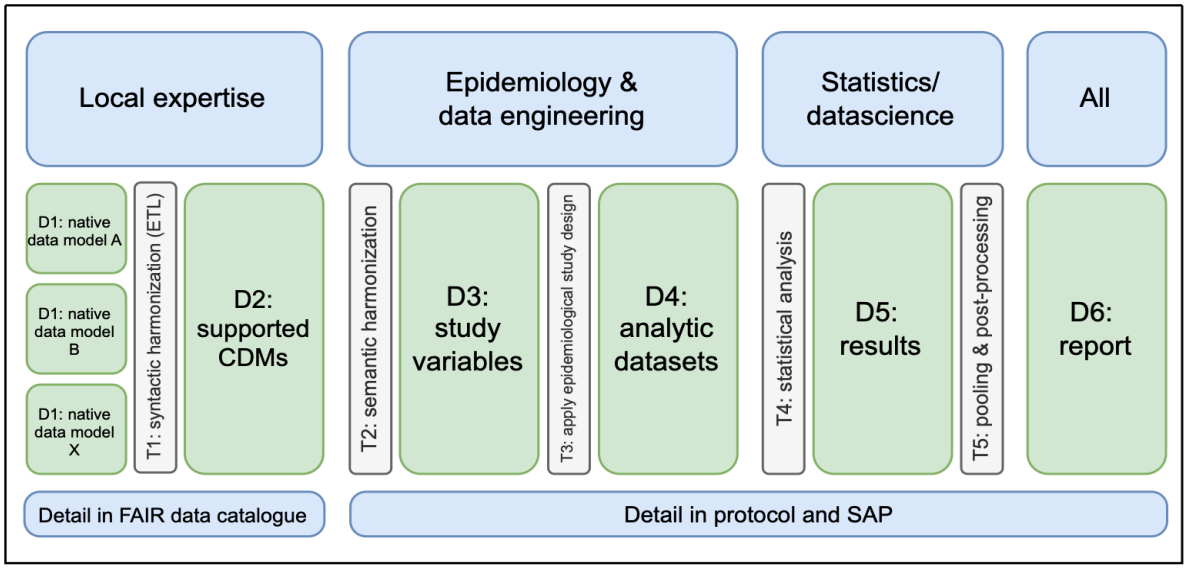


Figure 2. Data Management from the data transformation perspective

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 2). 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.

8.9. *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.

9. Deviations from Protocol

There were no deviations related to the estimand or design

Deviations related to the analysis

Frailty was excluded from the covariate list for both CPRD and BIFAP, as it could not be measured validly or consistently across the two data sources. Its omission from the propensity score model may have introduced residual confounding; however, because frailty is more likely to predict all-cause mortality than cardiovascular mortality specifically, the anticipated impact on this outcome is minimal.

Smoking status was excluded as a covariate from BIFAP because it was missing for 76.5% patients at baseline which exceeded our 40% threshold for acceptable levels of missingness for imputation.

For CPRD, missing smoking status was replaced as "none" given that recording smoking status, and encouraging smoking cessation for patients by primary care providers is financially incentivised in the UK; current and former smokers are also more likely to have their status reconfirmed at subsequent contacts as part of this cessation-advice pathway, whereas a "never smoker" status typically requires no further follow-up coding once recorded. This means status updates are particularly well recorded for current and former smokers, whilst those with missing smoking status are more likely to be never smokers. The degree to which missingness is at random for smoking status is debatable, and potentially represents an unreasonable assumption. There are examples of previous UK EHR based research where patients with missing status were assumed to be non-smokers: Williamson et al. made the same assumption in OpenSAFELY and reported that results were similar when missing data were instead handled via multiple imputation (ref https://www.nature.com/articles/s41586-020-2521-4). Furthermore, national survey data show the prevalence of never smoking to be ~63% amongst adults in England (https://www.gov.uk/government/statistics/smoking-profile-november-2025-update/smoking-profile-

statistical-commentary-november-2025). If we assume all missingness in our dataset to be from non-smokers, the prevalence of never smoking would be ~53%. Given the national survey findings, and the age and disease status of our study population, this seems reasonable, though as with any check against marginal prevalence, this confirms aggregate plausibility rather than correct assignment at the individual level.

Our original plan was to apply multiple imputation for HbA1c across the analyses for all three estimands. We subsequently revised this approach, restricting multiple imputation to a sensitivity analysis of the primary estimand. This sensitivity analysis was conducted using CPRD only, as missingness in BIFAP was too above our 40% threshold.

Tipping point analyses were planned to assess sensitivity to informative censoring, but were not warranted: non-administrative censoring was absent from the primary estimand in CPRD and accounted for only 3% of censoring in BIFAP.

10. Results

10.1. *Study population*

After application of the eligibility criteria, the final study cohorts comprised 34,110 patients in CPRD and 205,993 patients in BIFAP. In CPRD, 3,850 (11.3%) patients initiated dapagliflozin and 30,260 (88.7%) initiated a DPP4-i. In BIFAP, 71,159 (34.5%) patients initiated dapagliflozin and 134,834 initiated a DPP4-i (65.5%). (Figure 3 & Figure 4). This represents the study population across all estimands.

10.2. *Patient characteristics*

For BIFAP, the proportion of missing data for smoking was 76.5%, therefore, this variable was removed. For CPRD, it was missing for 0.3%.

For BIFAP, HbA1c was missing for 60.7% of patients and was therefore removed. In CPRD it was only missing in 3% of patients and therefore was imputed using MICE (in a sensitivity analysis).

Propensity score distributions by treatment group are shown for CPRD (Supplemental Figure 1) and BIFAP (Supplemental Figure 2). In both databases, the dapagliflozin and Dpp4-i groups showed substantial overlap across the range of estimated propensity scores, with the CPRD distributions more heavily right-skewed and concentrated at lower values and the BIFAP distributions more widely spread; in neither database did either group exhibit regions of zero density within the other's range, indicating that the positivity assumption was not violated. Given adequate overlap in both cohorts, no propensity score trimming or restriction was applied.

Before weighting, dapagliflozin initiators differed systematically from DPP4-i initiators in both databases (Table 19 & Table 20). In CPRD, dapagliflozin users were younger (mean age 70.6 vs. 75.1 years), more often male (68.5% vs. 57.3%; ASD = 0.2), and more frequently prescribed insulin (14.2% vs. 7.2%; ASD = 0.2) and aldosterone antagonists (19.4% vs. 6.4%; ASD = 0.4), while sulfonylurea use was less common (29.4% vs. 47.3%; ASD = 0.4). A similar direction of imbalance was seen in BIFAP, where dapagliflozin users were younger (71.7 vs. 76.1 years) and more often male (60.1% vs. 48.7%; ASD = 0.2), though comedication imbalances were generally smaller in magnitude than in CPRD (all pre-weighting ASDs ≤0.2). Baseline metformin use also differed markedly between databases (73–79% in CPRD vs. 39–48% in BIFAP). After weighting, covariate balance improved substantially in both databases. In BIFAP, all reported standardised differences were reduced to ≤0.05 (rounding to 0.0), indicating essentially complete balance across demographics, comorbidities and comedications. In CPRD, balance also improved markedly, with standardised differences ≤0.1 for nearly all covariates.

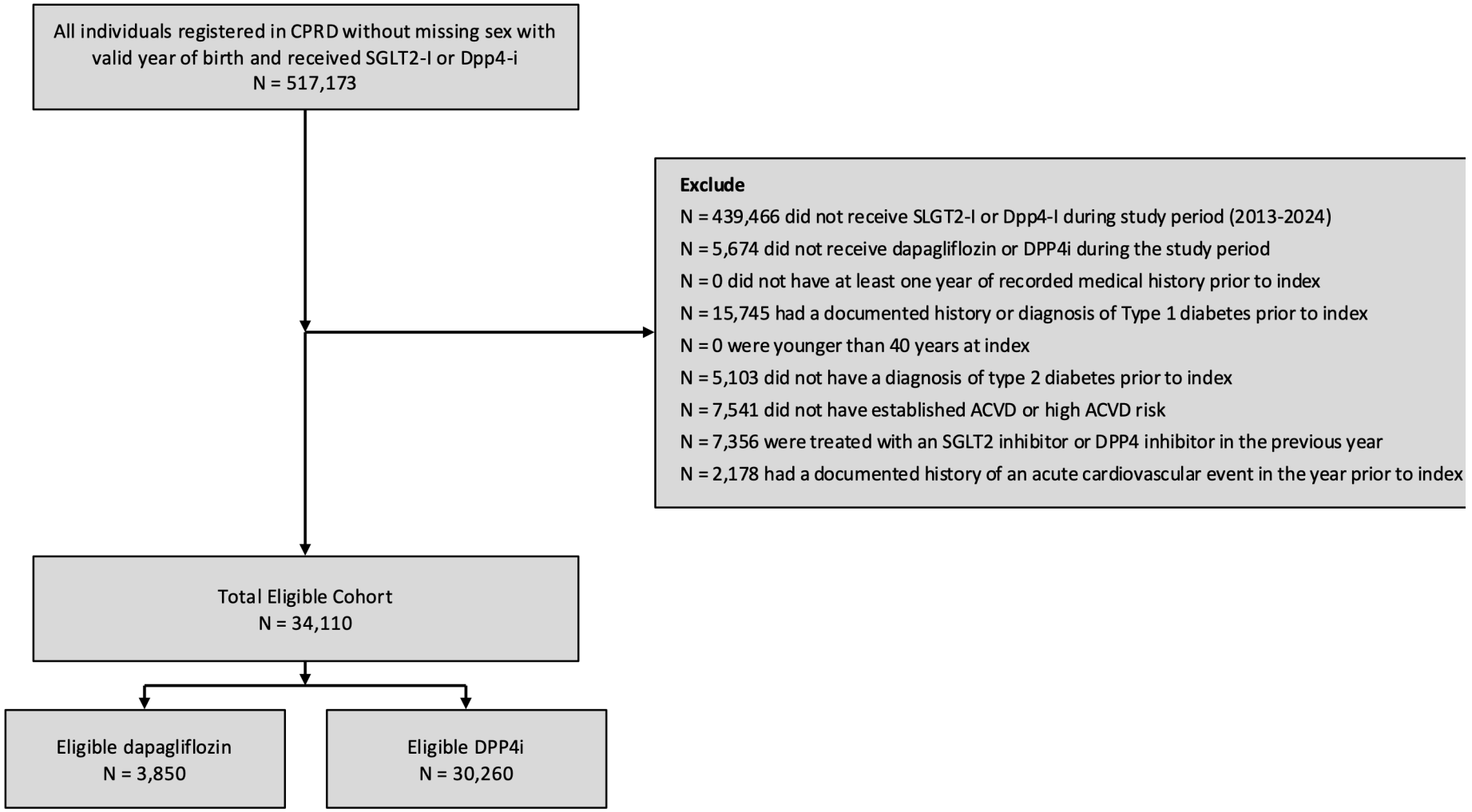


Figure 3. Flowchart of patient exclusions, CPRD

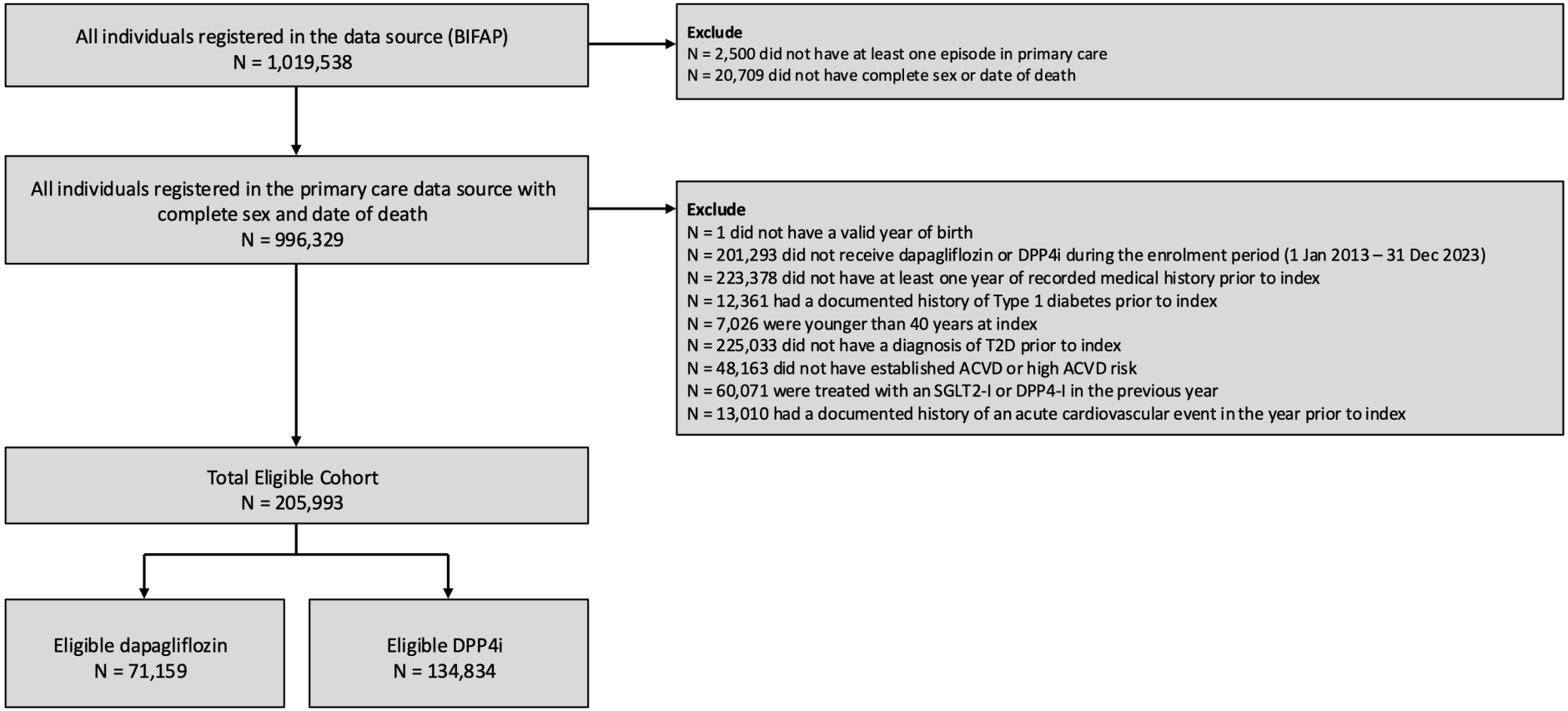


Figure 4. Flowchart of patient exclusions, BIFAP

Table 19. Baseline characteristics of the study population overall and by treatment group (dapagliflozin vs DPP-4 inhibitors), before and after IPTW, CPRD

	Before weighting				After weighting			
	Overall (n=34,110)	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**	Overall	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**
Demographics								
Age category N (%)								
<65	6,262 (18.4%)	1,266 (32.9%)	4,996 (16.5%)	0.3	6,321.3 (18.8%)	775.7 (22.9%)	5,545.6 (18.4%)	0.1
65-74	9,961 (29.2%)	1,234 (32.1%)	8,727 (28.8%)	0.1	10,020.5 (29.8%)	1,173.6 (34.7%)	8,846.9 (29.3%)	0.1
75-84	11,900 (34.9%)	869 (22.6%)	11,031 (36.5%)	0.3	11,536.6 (34.3%)	997.4 (29.5%)	10,539.2 (34.9%)	0.1
>=85	5,987 (17.6%)	481 (12.5%)	5,506 (18.2%)	0.2	5,722 (17%)	435.1 (12.9%)	5,286.9 (17.5%)	0.1
Years with diabetes								
Mean (SD)	8.8 (5.3)	8.4 (5.7)	8.9 (5.3)	0.1	8.8 (5.3)	8.7 (5.6)	8.8 (5.3)	0
Median (Q1 - Q3)	9 (5-13)	8 (4-12)	9 (5-13)	NA		8 (4-13)	9 (5-13)	NA
Gender N (%)								
Male	19,977 (58.6%)	2,637 (68.5%)	17,340 (57.3%)	0.2	19,760.6 (58.8%)	2,086.8 (61.7%)	17,673.8 (58.5%)	0.1
Female	14,132 (41.4%)	1,213 (31.5%)	12,919 (42.7%)	0.2	13,838.8 (41.2%)	1,294.9 (38.3%)	12,543.9 (41.5%)	0.1
Other	< 5	0 (0%)	< 5	0	0.9 (0%)	0 (0%)	0.9 (0%)	0
Comorbidities N (%)								

	Before weighting				After weighting			
	Overall (n=34,110)	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**	Overall	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**
Chronic Kidney Disease N (%)								
Stage 1-4	27,763 (81.4%)	3,108 (80.7%)	24,655 (81.5%)	0	27,349 (81.4%)	2,759.5 (81.6%)	24,589.5 (81.4%)	0
Stage 5	5,403 (15.8%)	571 (14.8%)	4,832 (16%)	0	5,296 (15.8%)	506.8 (15%)	4,789.2 (15.8%)	0
Obesity	2,253 (6.6%)	410 (10.6%)	1,843 (6.1%)	0.2	2,244 (6.7%)	265.2 (7.8%)	1,978.8 (6.5%)	0.1
Heart transplant	116 (0.3%)	29 (0.8%)	87 (0.3%)	0.1	114 (0.3%)	13.5 (0.4%)	100.5 (0.3%)	0
Microvascular complications	263 (0.8%)	34 (0.9%)	229 (0.8%)	0	252.4 (0.8%)	25.6 (0.8%)	226.8 (0.8%)	0
Severe hypoglycaemia	< 5	0 (0%)	< 5	0	0.9 (0%)	0 (0%)	0.9 (0%)	0
Diabetic neuropathy	1,497 (4.4%)	133 (3.5%)	1,364 (4.5%)	0.1	1,458.3 (4.3%)	130.5 (3.9%)	1,327.8 (4.4%)	0
Diabetic Retinopathy	10,203 (29.9%)	965 (25.1%)	9,238 (30.5%)	0.1	10,020.4 (29.8%)	964.7 (28.5%)	9,055.7 (30%)	0
Keto-/lactate acidosis	91 (0.3%)	5 (0.1%)	86 (0.3%)	0	84.6 (0.3%)	4 (0.1%)	80.6 (0.3%)	0
Lower limb amputations	359 (1.1%)	33 (0.9%)	326 (1.1%)	0	347 (1%)	31.1 (0.9%)	315.9 (1%)	0
COPD	5,854 (17.2%)	725 (18.8%)	5,129 (16.9%)	0	5,727.5 (17%)	560.1 (16.6%)	5,167.4 (17.1%)	0
Cancer	8,777 (25.7%)	933 (24.2%)	7,844 (25.9%)	0	8,615.7 (25.6%)	842.7 (24.9%)	7,773 (25.7%)	0
Major organ specific bleeding	643 (1.9%)	70 (1.8%)	573 (1.9%)	0	640.3 (1.9%)	70.2 (2.1%)	570.1 (1.9%)	0

	Before weighting				After weighting			
	Overall (n=34,110)	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**	Overall	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**
Bariatric surgery	30 (0.1%)	6 (0.2%)	24 (0.1%)	0	32.4 (0.1%)	5 (0.1%)	27.4 (0.1%)	0
Kidney Transplantation	120 (0.4%)	< 5	116 (0.4%)	0.1	113.8 (0.3%)	7.2 (0.2%)	106.6 (0.4%)	0
Dialysis	98 (0.3%)	< 5	95 (0.3%)	0.1	91.7 (0.3%)	4.7 (0.1%)	87 (0.3%)	0
Comedications N (%)								
Metformin	26,675 (78.2%)	2,835 (73.6%)	23,840 (78.8%)	0.1	26,382.4 (78.5%)	2,708.7 (80.1%)	23,673.7 (78.3%)	0
Sulfonylurea	15,430 (45.2%)	1,131 (29.4%)	14,299 (47.3%)	0.4	15,151.6 (45.1%)	1,433.6 (42.4%)	13,718 (45.4%)	0.1
Sulfonamides	3,124 (9.2%)	165 (4.3%)	2,959 (9.8%)	0.2	3,011.4 (9%)	236.6 (7%)	2,774.8 (9.2%)	0.1
Alfa glucosidase inhibitors	96 (0.3%)	6 (0.2%)	90 (0.3%)	0	91.6 (0.3%)	7 (0.2%)	84.6 (0.3%)	0
Thiazolidinediones	1,948 (5.7%)	187 (4.9%)	1,761 (5.8%)	0	1,935.4 (5.8%)	203.9 (6%)	1,731.5 (5.7%)	0
Other GLDs	140 (0.4%)	13 (0.3%)	127 (0.4%)	0	141.3 (0.4%)	16.7 (0.5%)	124.6 (0.4%)	0
Insulin	2,731 (8%)	548 (14.2%)	2,183 (7.2%)	0.2	2,721.7 (8.1%)	315 (9.3%)	2,406.7 (8%)	0
ACE inhibitors	16,798 (49.2%)	2,038 (52.9%)	14,760 (48.8%)	0.1	16,570.9 (49.3%)	1,702.3 (50.3%)	14,868.6 (49.2%)	0
ARBs	7,348 (21.5%)	818 (21.2%)	6,530 (21.6%)	0	7,205.1 (21.4%)	700 (20.7%)	6,505.1 (21.5%)	0

	Before weighting				After weighting			
	Overall (n=34,110)	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**	Overall	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**
Beta-blockers	13,334 (39.1%)	1,773 (46.1%)	11,561 (38.2%)	0.2	13,079 (38.9%)	1,318.5 (39%)	11,760.5 (38.9%)	0
Low/high ceiling diuretics	9,757 (28.6%)	1,367 (35.5%)	8,390 (27.7%)	0.2	9,559.5 (28.5%)	965.6 (28.6%)	8,593.9 (28.4%)	0
Aldosterone antagonists	2,674 (7.8%)	746 (19.4%)	1,928 (6.4%)	0.4	2,583 (7.7%)	291.1 (8.6%)	2,291.9 (7.6%)	0
Thiazide diuretics	3,891 (11.4%)	305 (7.9%)	3,586 (11.9%)	0.1	3,798.8 (11.3%)	348.3 (10.3%)	3,450.5 (11.4%)	0
Calcium channel blockers	13,371 (39.2%)	1,300 (33.8%)	12,071 (39.9%)	0.1	13,164.9 (39.2%)	1,292.5 (38.2%)	11,872.4 (39.3%)	0
Digitoxin/digoxin	2,407 (7.1%)	268 (7%)	2,139 (7.1%)	0	2,322.1 (6.9%)	200 (5.9%)	2,122.1 (7%)	0
Antiarrhythmics	452 (1.3%)	94 (2.4%)	358 (1.2%)	0.1	428.9 (1.3%)	46.3 (1.4%)	382.6 (1.3%)	0
Statins	27,391 (80.3%)	3,105 (80.6%)	24,286 (80.3%)	0	26,992.5 (80.3%)	2,727.4 (80.7%)	24,265.1 (80.3%)	0
Warfarin	3,364 (9.9%)	228 (5.9%)	3,136 (10.4%)	0.2	3,273.8 (9.7%)	280.3 (8.3%)	2,993.5 (9.9%)	0.1
Aspirin	11,865 (34.8%)	1,171 (30.4%)	10,694 (35.3%)	0.1	11,647.9 (34.7%)	1,128.1 (33.4%)	10,519.8 (34.8%)	0
Purinergic receptor $P2Y_{12}$ antagonists	3,082 (9%)	348 (9%)	2,734 (9%)	0	3,026.9 (9%)	306.2 (9.1%)	2,720.7 (9%)	0
Other antiplatelets	11,869 (34.8%)	1,170 (30.4%)	10,699 (35.4%)	0.1	11,653 (34.7%)	1,129.6 (33.4%)	10,523.4 (34.8%)	0

	Before weighting				After weighting			
	Overall (n=34,110)	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**	Overall	Dapagliflozin (n=3,850)	Dpp4-i (n=30,260)	ASD**
Corticosteroids	5,714 (16.8%)	656 (17%)	5,058 (16.7%)	0	5,608.9 (16.7%)	556.7 (16.5%)	5,052.2 (16.7%)	0
Weight loss drugs	216 (0.6%)	41 (1.1%)	175 (0.6%)	0.1	220.6 (0.7%)	28.2 (0.8%)	192.4 (0.6%)	0
Lifestyle N (%)								
Tobacco use**								
None	9,978 (29.3%)	1,036 (26.9%)	9,043 (29.9%)	0.1	9,919.5 (29.5%)	981.4 (29%)	8,938.1 (29.6%)	0
Past	20,438 (59.9%)	2,340 (60.8%)	18,098 (59.8%)	0	20,111.8 (59.9%)	2,018.6 (59.7%)	18,093.2 (59.9%)	0
Current	3,593 (10.5%)	474 (12.3%)	3,119 (10.3%)	0.1	3,568.9 (10.6%)	381.7 (11.3%)	3,187.2 (10.5%)	0
Established ACVD								
Ischemic heart disease	4,999 (14.7%)	719 (18.7%)	4,280 (14.1%)	0.1	4,871.3 (14.5%)	479.3 (14.2%)	4,392 (14.5%)	0
Ischemic cerebrovascular disease	629 (1.8%)	56 (1.5%)	573 (1.9%)	0	602.9 (1.8%)	47.2 (1.4%)	555.7 (1.8%)	0
ACVD Risk factors								
Peripheral arterial disease	339 (1%)	57 (1.5%)	282 (0.9%)	0.1	320.6 (1%)	31.3 (0.9%)	289.3 (1%)	0
Dyslipidaemia	3,172 (9.3%)	378 (9.8%)	2,794 (9.2%)	0	3,117 (9.3%)	308.7 (9.1%)	2,808.3 (9.3%)	0
Hypertension	14,723 (43.2%)	1,715 (44.5%)	13,008 (43%)	0	14,414.7 (42.9%)	1,410.9 (41.7%)	13,003.8 (43%)	0

**ASD: For continuous variables, calculated as the absolute difference in group means divided by the pooled standard deviation, where the pooled standard deviation is the square root of the average of the two groups' variances. For categorical variables, each level is treated as a binary indicator and the ASD is calculated as the absolute difference in proportions divided by the standard deviation of the two proportions, with each proportion's variance calculated as p(1−p); one ASD value is reported per level.

Table 20. Baseline characteristics of the study population overall and by treatment group (dapagliflozin vs DPP-4 inhibitors), before and after IPTW, BIFAP

	Before trimming				After trimming			
	Overall (n=205,993)	Dapagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**	Overall	DApagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**
Demographics								
Age category N (%)								
<65	39,785 (19.3%)	18,563 (26.1%)	21,222 (15.7%)	0.2	40,142.6 (19.6%)	13,885.2 (19.9%)	26,257.4 (19.5%)	0
65-74	61,676 (29.9%)	24,883 (35%)	36,793 (27.3%)	0.2	61,718.5 (30.2%)	21,276.9 (30.5%)	40,441.6 (30%)	0
75-84	63,597 (30.9%)	19,796 (27.8%)	43,801 (32.5%)	0.1	63,070.3 (30.8%)	21,539.6 (30.9%)	41,530.7 (30.8%)	0
>=85	40,935 (19.9%)	7,917 (11.1%)	33,018 (24.5%)	0.4	39,735.3 (19.4%)	12,971.2 (18.6%)	26,764.1 (19.8%)	0
Years with diabetes								
Mean (SD)	4.1 (3.7)	4.7 (3.8)	3.8 (3.6)	0.2	4.1 (3.7)	4.2 (3.7)	4.1 (3.7)	0
Median (Q1 - Q3)	3 (1-7)	4 (1-7)	3 (1-6)	NA		4 (1-7)	3 (1-7)	NA
Gender N (%)								
Male	108,518 (52.7%)	42,787 (60.1%)	65,731 (48.7%)	0.2	108,358.2 (52.9%)	37,147 (53.3%)	71,211.2 (52.8%)	0
Female	97,475 (47.3%)	28,372 (39.9%)	69,103 (51.3%)	0.2	96,308.6 (47.1%)	32,525.9 (46.7%)	63,782.7 (47.2%)	0
Other	0 (0%)	0 (0%)	0 (0%)	0	0 (0%)	0 (0%)	0 (0%)	0

	Before trimming				After trimming			
	Overall (n=205,993)	Dapagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**	Overall	DApagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**
Comorbidities N (%)								
Chronic Kidney Disease N (%)								
Stage 1-4	90,746 (44.1%)	28,067 (39.4%)	62,679 (46.5%)	0.2	89,612 (43.8%)	30,296.2 (43.5%)	59,315.8 (43.9%)	0
Stage 5	16,345 (7.9%)	3,829 (5.4%)	12,516 (9.3%)	0.2	15,938.4 (7.8%)	5,211.9 (7.5%)	10,726.5 (7.9%)	0
Obesity	13,080 (6.3%)	5,046 (7.1%)	8,034 (6%)	0	13,018.1 (6.4%)	4,466.7 (6.4%)	8,551.4 (6.3%)	0
Heart transplant	163 (0.1%)	72 (0.1%)	91 (0.1%)	0	178 (0.1%)	63.7 (0.1%)	114.3 (0.1%)	0
Microvascular complications	4,060 (2%)	1,695 (2.4%)	2,365 (1.8%)	0	4,055.4 (2%)	1,390.3 (2%)	2,665.1 (2%)	0
Severe hypoglycaemia	63 (0%)	32 (0%)	31 (0%)	0	63 (0%)	21.8 (0%)	41.2 (0%)	0
Diabetic neuropathy	5,373 (2.6%)	1,968 (2.8%)	3,405 (2.5%)	0	5,448.2 (2.7%)	1,874.8 (2.7%)	3,573.4 (2.6%)	0
Diabetic Retinopathy	8,474 (4.1%)	3,103 (4.4%)	5,371 (4%)	0	8,502.7 (4.2%)	2,891.5 (4.2%)	5,611.2 (4.2%)	0
Keto-/lactate acidosis	477 (0.2%)	121 (0.2%)	356 (0.3%)	0	469.6 (0.2%)	155.8 (0.2%)	313.8 (0.2%)	0
Lower limb amputations	1,215 (0.6%)	408 (0.6%)	807 (0.6%)	0	1,206.3 (0.6%)	409.1 (0.6%)	797.2 (0.6%)	0
COPD	22,600 (11%)	8,104 (11.4%)	14,496 (10.8%)	0	22,573.1 (11%)	7,750.1 (11.1%)	14,823 (11%)	0
Cancer	36,954 (17.9%)	11,160 (15.7%)	25,794 (19.1%)	0.1	36,781.5 (18%)	12,523.7 (18%)	24,257.8 (18%)	0
Major organ specific bleeding	2,602 (1.3%)	577 (0.8%)	2,025 (1.5%)	0.1	2,567.5 (1.3%)	840.9 (1.2%)	1,726.6 (1.3%)	0

	Before trimming				After trimming			
	Overall (n=205,993)	Dapagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**	Overall	DApagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**
Bariatric surgery	353 (0.2%)	184 (0.3%)	169 (0.1%)	0	358.8 (0.2%)	126.1 (0.2%)	232.7 (0.2%)	0
Kidney Transplantation	984 (0.5%)	219 (0.3%)	765 (0.6%)	0	938.5 (0.5%)	291.4 (0.4%)	647.1 (0.5%)	0
Dialysis	655 (0.3%)	40 (0.1%)	615 (0.5%)	0.1	519.4 (0.3%)	90.3 (0.1%)	429.1 (0.3%)	0
Comedications N (%)								
Metformin	92,729 (45%)	27,587 (38.8%)	65,142 (48.3%)	0.2	91,432.5 (44.7%)	30,895.1 (44.3%)	60,537.4 (44.8%)	0
Sulfonylurea	18,357 (8.9%)	5,703 (8%)	12,654 (9.4%)	0	17,908.1 (8.7%)	5,951.7 (8.5%)	11,956.4 (8.9%)	0
Sulfonamides	4,730 (2.3%)	1,077 (1.5%)	3,653 (2.7%)	0.1	4,587.8 (2.2%)	1,479.8 (2.1%)	3,108 (2.3%)	0
Alfa glucosidase inhibitors	598 (0.3%)	119 (0.2%)	479 (0.4%)	0	560.6 (0.3%)	171.9 (0.2%)	388.7 (0.3%)	0
Thiazolidinediones	783 (0.4%)	312 (0.4%)	471 (0.3%)	0	777 (0.4%)	264.9 (0.4%)	512.1 (0.4%)	0
Other GLDs	16,894 (8.2%)	5,400 (7.6%)	11,494 (8.5%)	0	16,653.5 (8.1%)	5,591.2 (8%)	11,062.3 (8.2%)	0
Insulin	36,310 (17.6%)	12,400 (17.4%)	23,910 (17.7%)	0	36,069.3 (17.6%)	12,208.7 (17.5%)	23,860.6 (17.7%)	0
ACE inhibitors	46,379 (22.5%)	15,858 (22.3%)	30,521 (22.6%)	0	46,059.4 (22.5%)	15,642.1 (22.5%)	30,417.3 (22.5%)	0
ARBs	42,493 (20.6%)	14,878 (20.9%)	27,615 (20.5%)	0	42,455.8 (20.7%)	14,544.5 (20.9%)	27,911.3 (20.7%)	0
Beta-blockers	66,737 (32.4%)	25,229 (35.5%)	41,508 (30.8%)	0.1	67,064.4 (32.8%)	23,132 (33.2%)	43,932.4 (32.5%)	0

	Before trimming				After trimming			
	Overall (n=205,993)	Dapagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**	Overall	DApagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**
Low/high ceiling diuretics	61,318 (29.8%)	19,614 (27.6%)	41,704 (30.9%)	0.1	60,948.3 (29.8%)	20,754 (29.8%)	40,194.3 (29.8%)	0
Aldosterone antagonists	19,145 (9.3%)	8,326 (11.7%)	10,819 (8%)	0.1	19,240.3 (9.4%)	6,698.6 (9.6%)	12,541.7 (9.3%)	0
Thiazide diuretics	6,204 (3%)	1,983 (2.8%)	4,221 (3.1%)	0	6,150.5 (3%)	2,083.4 (3%)	4,067.1 (3%)	0
Calcium channel blockers	48,190 (23.4%)	15,440 (21.7%)	32,750 (24.3%)	0.1	47,747.3 (23.3%)	16,149.7 (23.2%)	31,597.6 (23.4%)	0
Digitoxin/digoxin	7,048 (3.4%)	2,036 (2.9%)	5,012 (3.7%)	0	6,982.9 (3.4%)	2,350.9 (3.4%)	4,632 (3.4%)	0
Antiarrhythmics	5,533 (2.7%)	2,236 (3.1%)	3,297 (2.4%)	0	5,544.2 (2.7%)	1,917.5 (2.8%)	3,626.7 (2.7%)	0
Statins	125,168 (60.8%)	44,918 (63.1%)	80,250 (59.5%)	0.1	124,857.3 (61%)	42,700.9 (61.3%)	82,156.4 (60.9%)	0
Warfarin	910 (0.4%)	294 (0.4%)	616 (0.5%)	0	895.8 (0.4%)	299.9 (0.4%)	595.9 (0.4%)	0
Aspirin	59,089 (28.7%)	20,588 (28.9%)	38,501 (28.6%)	0	58,566 (28.6%)	19,825.5 (28.5%)	38,740.5 (28.7%)	0
Purinergic receptor $P2Y_{12}$ antagonists	10,779 (5.2%)	3,621 (5.1%)	7,158 (5.3%)	0	10,739.8 (5.2%)	3,645.4 (5.2%)	7,094.4 (5.3%)	0
Other antiplatelets	60,393 (29.3%)	20,987 (29.5%)	39,406 (29.2%)	0	59,851.3 (29.2%)	20,256.5 (29.1%)	39,594.8 (29.3%)	0
Corticosteroids	30,705 (14.9%)	9,815 (13.8%)	20,890 (15.5%)	0	30,536.5 (14.9%)	10,375.6 (14.9%)	20,160.9 (14.9%)	0
Weight loss drugs	8 (0%)	< 5	5 (0%)	0	7.9 (0%)	2.4 (0%)	5.5 (0%)	0
Lifestyle N (%)								

	Before trimming				After trimming			
	Overall (n=205,993)	Dapagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**	Overall	DApagliflozin (n=71,159)	Dpp4-i (n=134,834)	ASD**
Tobacco use								
None	157,811 (76.6%)	51,879 (72.9%)	105,932 (78.6%)	0.1	156,326 (76.4%)	53,027.1 (76.1%)	103,298.9 (76.5%)	0
Past	22,635 (11%)	8,550 (12%)	14,085 (10.4%)	0	22,658.4 (11.1%)	7,779.6 (11.2%)	14,878.8 (11%)	0
Current	25,547 (12.4%)	10,730 (15.1%)	14,817 (11%)	0.1	25,682.2 (12.5%)	8,866.1 (12.7%)	16,816.1 (12.5%)	0
Established ACVD								
Ischemic heart disease	14,114 (6.9%)	4,969 (7%)	9,145 (6.8%)	0	14,101.6 (6.9%)	4,818.6 (6.9%)	9,283 (6.9%)	0
Ischemic cerebrovascular disease	7,814 (3.8%)	1,808 (2.5%)	6,006 (4.5%)	0.1	7,554.5 (3.7%)	2,439.3 (3.5%)	5,115.2 (3.8%)	0
ACVD Risk factors								
Peripheral arterial disease	4,052 (2%)	1,170 (1.6%)	2,882 (2.1%)	0	3,968.6 (1.9%)	1,310.7 (1.9%)	2,657.9 (2%)	0
Dyslipidaemia	41,808 (20.3%)	13,363 (18.8%)	28,445 (21.1%)	0.1	41,570.9 (20.3%)	14,085.1 (20.2%)	27,485.8 (20.4%)	0
Hypertension	56,237 (27.3%)	16,454 (23.1%)	39,783 (29.5%)	0.1	55,319.2 (27%)	18,487.2 (26.5%)	36,832 (27.3%)	0

*ASD: For continuous variables, calculated as the absolute difference in group means divided by the pooled standard deviation, where the pooled standard deviation is the square root of the average of the two groups' variances. For categorical variables, each level is treated as a binary indicator and the ASD is calculated as the absolute difference in proportions divided by the standard deviation of the two proportions, with each proportion's variance calculated as p(1−p); one ASD value is reported per level.

10.3. *Outcome Descriptive Summary Statistics*

Under estimand 1, median follow-up was consistently shorter for dapagliflozin than DPP4-i in both databases (

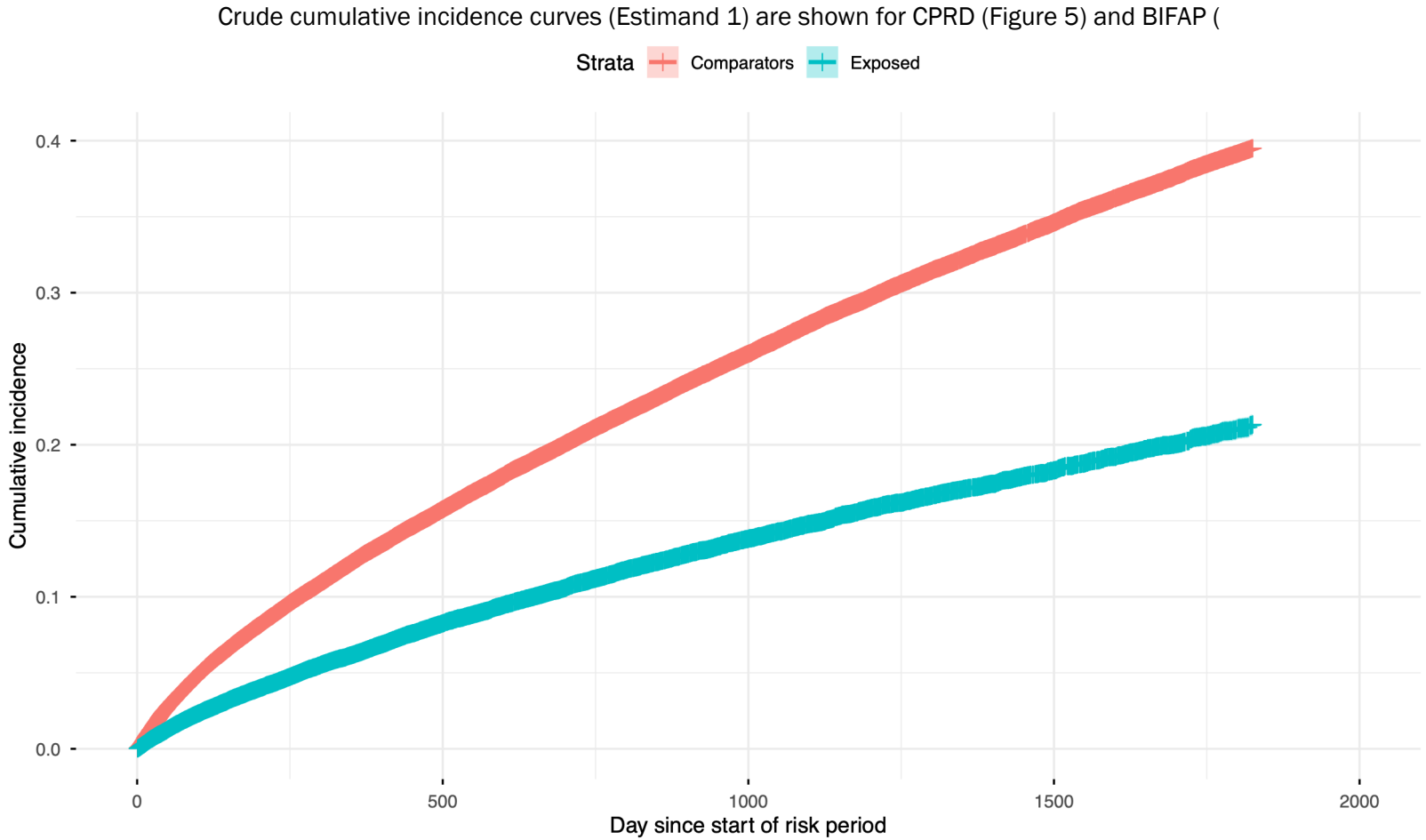
Table 21 &

Table 22): 16.1 months (Q1–Q3: 6.7–35.5) versus 33.3 months (14.2–59.5) in CPRD, and 13.9 months (6.7–26.7) versus 23.2 months (9.4–42.9) in BIFAP.

Reasons for end of follow-up differed markedly between databases. In CPRD, follow-up most often ended in the outcome itself (Estimand 1: 47.6% DPP4-i, 45.0% dapagliflozin) or end of observation in the data source (28.3%, 42.5%), with the remainder censored at the fixed administrative cut-off (24.1%, 12.5% at 5 years; no patients censored at the overall end of the study period). For estimand 2, treatment discontinuation was introduced as an additional censoring reason (end of index episode: 18.7% DPP4-i, 20.5% dapagliflozin; switch to comparator: 33.1%, 16.6%), and Estimand 3 mirrored Estimand 1 but with a 3-year rather than 5-year cut-off. In BIFAP, the pattern was reversed throughout: end of the study period was the dominant censoring reason under Estimand 1 (59.1% DPP4-i, 85.2% dapagliflozin), Estimand 2 (33.8%, 61.6%) and Estimand 3 (45.3%, 76.2%), while the outcome itself accounted for a comparatively small share (Estimand 1: 24.7%, 8.9%).

Under both Estimand 1 and Estimand 2, the direction of the treatment contrast differed between databases and was stable across estimands (

Table 24-Table 25). In CPRD, the crude incidence rate of MACE was higher for dapagliflozin than DPP4-i under Estimand 1 (237.2 vs. 169.8 per 1,000 person-years; 95% CI 226.2–248.7 vs. 167.1–172.6) and Estimand 2, where the gap widened (247.8 vs. 150.7; 95% CI 233.7–262.5 vs. 147.3–154.1). In BIFAP, the pattern was reversed and essentially unchanged between estimands: dapagliflozin showed a lower rate than DPP4-i under Estimand 1 (57.1 vs. 110.3; 95% CI 55.7–58.5 vs. 109.1–111.5) and Estimand 2 (56.5 vs. 107.9; 95% CI 54.9–58.2 vs. 106.4–109.4), with non-overlapping confidence intervals throughout.



*Exposed is dapagliflozin, Comparator is Dpp4-i

Figure 6) below. These visually confirm the incidence-rate findings above.

Absolute incidence rates were substantially higher in CPRD than BIFAP in both treatment arms (dapagliflozin 237.2 vs. 57.1 per 1,000 person-years; DPP4-i 169.8 vs. 110.3). The composition of the composite outcome also differed: death accounted for the large majority of events in CPRD (84–93%

across arms), whereas in BIFAP death predominated in the DPP4-i arm (84%) but comprised only about half of events in the dapagliflozin arm (55%), with AMI and stroke each contributing a further ~32%.

Table 21. Intercurrent events, and reasons for end of follow-up by treatment group (dapagliflozin vs DPP-4 inhibitors) for MACE, per estimand, CPRD

	Total	Dapagliflozin	DPP4-i
N	34,110	3,850	30,260
Estimand 1			
Cause of end of follow-up and time in person-months			
Outcome	16,134 (47.3%)	1,731 (45%)	14,403 (47.6%)
Time to occurrence, median (Q1–Q3)	22.1 (9.4–38.5)	12 (5.1–23.6)	23.6 (10.4–39.7)
Day 1825 following index date	7,783 (22.8%)	481 (12.5%)	7,302 (24.1%)
Time to occurrence, median (Q1–Q3)	60.8 (60.8–60.8)	60.8 (60.8–60.8)	60.8 (60.8–60.8)
End of observation of the person in the data source	10,193 (29.9%)	1,638 (42.5%)	8,555 (28.3%)
Time to occurrence, median (Q1–Q3)	21.6 (9.4–38)	14.7 (6.7–27.6)	23.4 (10.2–39.7)
Intercurrent events that do not cause end of follow-up and time in person-months			
Discontinuation	10,150 (29.8%)	1,116 (29%)	9,034 (29.9%)
Time to occurrence, median (Q1–Q3)	9.9 (4.9–23)	7.8 (4.9–17)	10.5 (4.9–23.8)
Switch/add on	12,516 (36.7%)	887 (23%)	11,629 (38.4%)
Time to occurrence, median (Q1–Q3)	11.9 (4.4–24.4)	10.6 (4.1–22.4)	12 (4.4–24.5)
Estimand 2			
Cause of end of follow-up			
Outcome	8,764 (25.7%)	1,155 (30%)	7,609 (25.1%)
Time to occurrence, median (Q1–Q3)	14.4 (5.4–29.3)	8.5 (3.5–17.2)	15.7 (6–31.3)
Day 1825 following index date	2,416 (7.1%)	139 (3.6%)	2,277 (7.5%)
Time to occurrence, median (Q1–Q3)	60.8 (60.8–60.8)	60.8 (60.8–60.8)	60.8 (60.8–60.8)
End of observation of the person in the data source	4712 (15.6%)	1125 (29.2%)	5837 (17.1%)
Time to occurrence, median (Q1–Q3)	15.9 (6.3-31.4)	11.7 (5-22)	14.9 (6-29.6)
Discontinuation	6,444 (18.9%)	790 (20.5%)	5,654 (18.7%)
Time to occurrence, median (Q1–Q3)	7.8 (4.9–18.4)	7.2 (4.9–15)	7.9 (4.9–19)
Switch/add on	10,649 (31.2%)	641 (16.6%)	10,008 (33.1%)

	Total	Dapagliflozin	DPP4-i
Time to occurrence, median (Q1–Q3)	10 (3.6–21.9)	6.6 (2.6–16.4)	10.3 (3.7–22.3)
End of observation of the person in the data source	5,837 (17.1%)	1,125 (29.2%)	4,712 (15.6%)
Time to occurrence, median (Q1–Q3)	14.9 (6–29.6)	11.7 (5–22)	15.9 (6.3–31.4)
Intercurrent events that do not cause end of follow-up			
None			
Estimand 3##			
Cause of end of follow-up			
Outcome	11,680 (34.2%)	1,516 (39.4%)	10,164 (33.6%)
Time to occurrence, median (Q1–Q3)	14.5 (6.5–24.5)	10.2 (4.5–18.3)	15.3 (7–25.2)
Day 1095 following index date	14,985 (43.9%)	929 (24.1%)	14,056 (46.5%)
Time to occurrence, median (Q1–Q3)	36.5 (36.5–36.5)	36.5 (36.5–36.5)	36.5 (36.5–36.5)
End of observation of the person in the data source	7,445 (21.8%)	1,405 (36.5%)	6,040 (20%)
Time to occurrence, median (Q1–Q3)	14.8 (6.8–24.5)	12.3 (5.8–21.2)	15.4 (7–25.1)
Intercurrent events that do not cause end of follow-up			
Discontinuation	8,913 (26.1%)	1,040 (27%)	7,873 (26%)
Time to occurrence, median (Q1–Q3)	8.4 (4.9–17)	7.2 (4.9–14.5)	8.6 (4.9–17.3)
Switch/add on	11,001 (32.3%)	798 (20.7%)	10,203 (33.7%)
Time to occurrence, median (Q1–Q3)	9.8 (3.7–19.1)	8.9 (3.6–18.1)	9.8 (3.7–19.2)

*Estimand 1: Treatment discontinuation, switching, and add-on antihyperglycemic therapy are handled via a treatment policy strategy
**Estimand 2: Treatment discontinuation, switching, and add-on antihyperglycemic therapy are handled via while on treatment strategy
#Estimand 3: Treatment discontinuation, switching, and add-on antihyperglycemic therapy are handled via a treatment policy strategy
Only 3 yr restricted FU is presented

Note for end of observation of the person in the data source: May include censoring at the practice's last date of data collection; administrative and non-administrative censoring could not be separated in this output.

Table 22. Intercurrent events, and reasons for end of follow-up by treatment group (dapagliflozin vs DPP-4 inhibitors) for MACE, per estimand, BIFAP

	Total	Dapagliflozin	DPP4-i
N	205,993	71,159	134,834
Estimand 1			
Cause of end of follow-up			
Outcome	39,588 (19.2%)	6,349 (8.9%)	33,239 (24.7%)
Time to occurrence, median (Q1–Q3)	12.8 (4.5–26.7)	9.2 (3.4–19.6)	13.7 (4.8–28.1)
Day 1825 following index date	20,475 (9.9%)	3,410 (4.8%)	17,065 (12.7%)
Time to occurrence, median (Q1–Q3)	60.8 (60.8–60.8)	60.8 (60.8–60.8)	60.8 (60.8–60.8)
End of the study	140,296 (68.1%)	60,636 (85.2%)	79,660 (59.1%)
Time to occurrence, median (Q1–Q3)	17.9 (8–31.6)	13.5 (6.8–25.1)	21.8 (9.7–35.3)
End of observation of the person in the data source	5,634 (2.7%)	764 (1.1%)	4,870 (3.6%)
Time to occurrence, median (Q1–Q3)	21.5 (9.6–39)	15.8 (7.3–31)	22.5 (10.1–40.3)
Intercurrent events that do not cause end of follow-up			
Discontinuation	56,516 (27.4%)	16,156 (22.7%)	40,360 (29.9%)
Time to occurrence, median (Q1–Q3)	8.6 (4.9–17.7)	7.3 (4.9–14)	9 (4.9–19.4)
Switch/add on	50,922 (24.7%)	11,065 (15.5%)	39,857 (29.6%)
Time to occurrence, median (Q1–Q3)	7.9 (1.9–20.1)	8.4 (2.8–19.2)	7.7 (1.7–20.4)
Estimand 2			
Cause of end of follow-up			
Outcome	24,404 (11.8%)	4,518 (6.3%)	19,886 (14.7%)
Time to occurrence, median (Q1–Q3)	7.9 (2.6–18.9)	6.5 (2.3–14.5)	8.3 (2.7–20.1)
Day 1825 following index date	6,606 (3.2%)	1,177 (1.7%)	5,429 (4%)
Time to occurrence, median (Q1–Q3)	60.8 (60.8–60.8)	60.8 (60.8–60.8)	60.8 (60.8–60.8)
End of the study	89,369 (43.4%)	43,851 (61.6%)	45,518 (33.8%)
Time to occurrence, median (Q1–Q3)	12.7 (5.9–25.1)	10.6 (5.3–20.6)	15.5 (6.6–29.8)
Discontinuation	40,141 (19.5%)	12,366 (17.4%)	27,775 (20.6%)

	Total	Dapagliflozin	DPP4-i
Time to occurrence, median (Q1–Q3)	7.5 (4.9–15.1)	6.7 (3.9–12.4)	7.8 (4.9–16.3)
Switch/add on	43,143 (20.9%)	8,881 (12.5%)	34,262 (25.4%)
Time to occurrence, median (Q1–Q3)	5.7 (1.4–16)	6 (1.9–14.9)	5.6 (1.3–16.3)
End of observation of the person in the data source	2,317 (1.1%)	364 (0.5%)	1,953 (1.4%)
Time to occurrence, median (Q1–Q3)	13.1 (4.1–28.3)	11.4 (3.3–21.2)	13.6 (4.4–29.7)
Intercurrent events that do not cause end of follow-up			
None			
Estimand 3##			
Cause of end of follow-up			
Outcome	34,071 (16.5%)	5,846 (8.2%)	28,225 (20.9%)
Time to occurrence, median (Q1–Q3)	10.1 (3.7–19.9)	8.2 (3.1–16.4)	10.5 (3.8–20.5)
Day 1095 following index date	52,628 (25.5%)	10,457 (14.7%)	42,171 (31.3%)
Time to occurrence, median (Q1–Q3)	36.5 (36.5–36.5)	36.5 (36.5–36.5)	36.5 (36.5–36.5)
End of the study	115,263 (56%)	54,234 (76.2%)	61,029 (45.3%)
Time to occurrence, median (Q1–Q3)	13.6 (6.8–23.3)	11.6 (6.2–20.8)	15.8 (7.6–25.8)
End of observation of the person in the data source	4,031 (2%)	622 (0.9%)	3,409 (2.5%)
Time to occurrence, median (Q1–Q3)	14.3 (6.5–23.7)	12.7 (5.8–21)	14.6 (6.7–24.2)
Intercurrent events that do not cause end of follow-up			
Discontinuation	52,598 (25.5%)	15,495 (21.8%)	37,103 (27.5%)
Time to occurrence, median (Q1–Q3)	7.7 (4.9–15.1)	6.8 (4.8–12.6)	8.1 (4.9–16)
Switch/add on	46,403 (22.5%)	10,213 (14.4%)	36,190 (26.8%)
Time to occurrence, median (Q1–Q3)	6.5 (1.6–16)	7.2 (2.4–16.1)	6.2 (1.4–16)

*Estimand 1: Treatment discontinuation, switching, and add-on antihyperglycemic therapy are handled via a treatment policy strategy
**Estimand 2: Treatment discontinuation, switching, and add-on antihyperglycemic therapy are handled via while on treatment strategy
#Estimand 3: Treatment discontinuation, switching, and add-on antihyperglycemic therapy are handled via a treatment policy strategy
Only 3 yr restricted FU is presented

Table 23. Crude incidence rates (per estimand) and restricted mean survival time for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors), CPRD

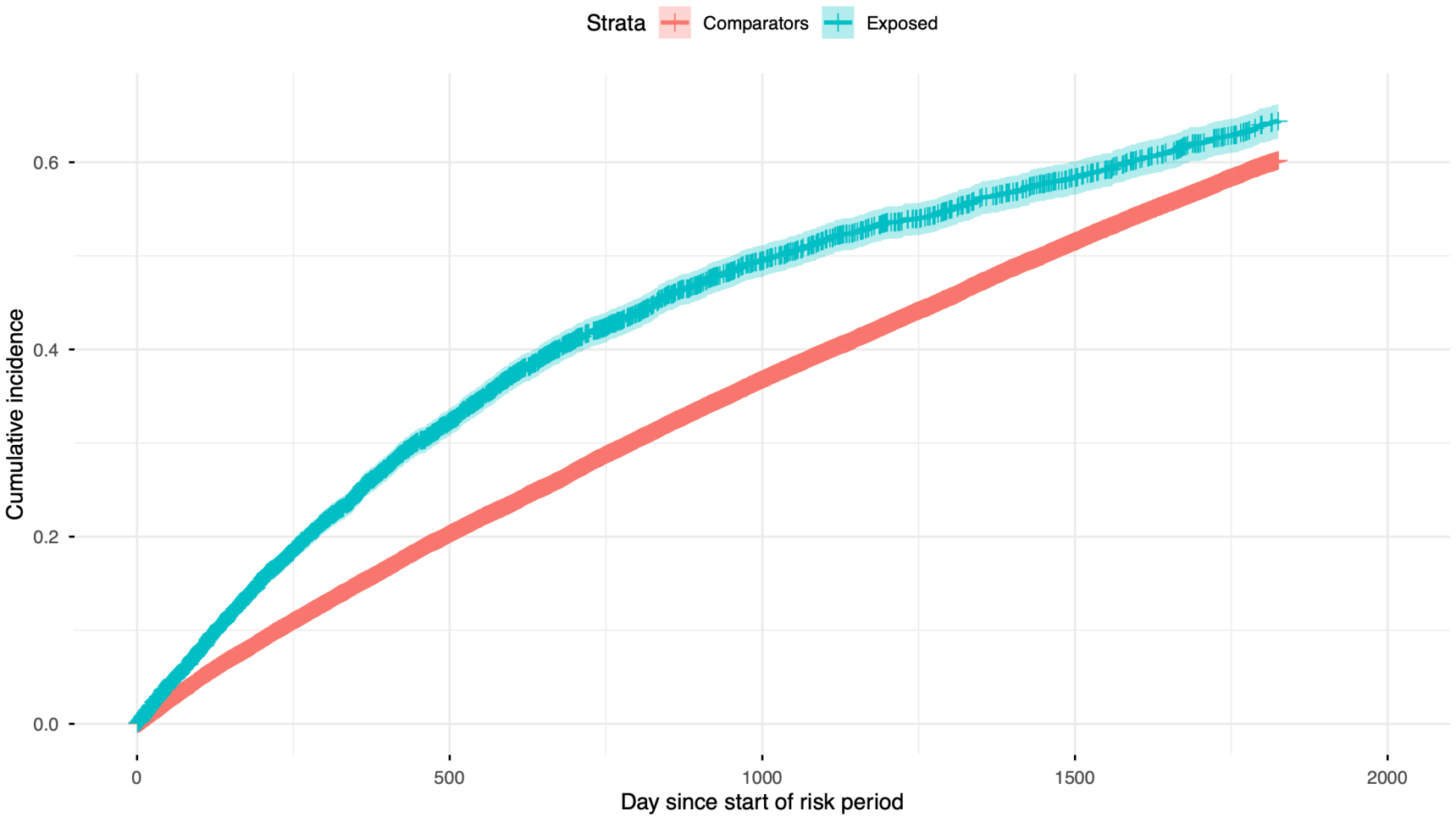
Comparison	Dapagliflozin						DPP4-I					
	N patients	N events	Person-years*	Incidence rate (95% CI)	RMST year 3 (95% CI)	RMST year 5 (95% CI)	N patients	N events	Person-years	Incidence rate (95% CI)	RMST year 3 (95% CI)	RMST year 5 (95% CI)
Estimand 1	3850	1731	7296.65	237.23 (226.2–248.7)	N/A	N/A	30260	14403	84806.57	169.83 (167.1–172.6)	N/A	N/A
Estimand 2	3850	1155	4661.07	247.8 (233.7–262.5)	N/A	N/A	30260	7609	50499.61	150.67 (147.3–154.1)	N/A	N/A
Estimand 3	3850	1516	5926.53	N/A	2.09 (2.06–2.13)	2.93 (2.86–3)	30260	10164	63879.25	N/A	2.39 (2.37–2.4)	3.38 (3.36–3.4)

*Person years: total follow-up time accumulated across all patients
Note: For Estimands 1 and 2, the number of events and accumulated person-time are reported over the maximum 5-year follow-up, whereas for Estimand 3 these quantities are reported over 3 years of follow-up only.

Table 24. Crude incidence rates (per estimand) and restricted mean survival time for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors), BIFAP

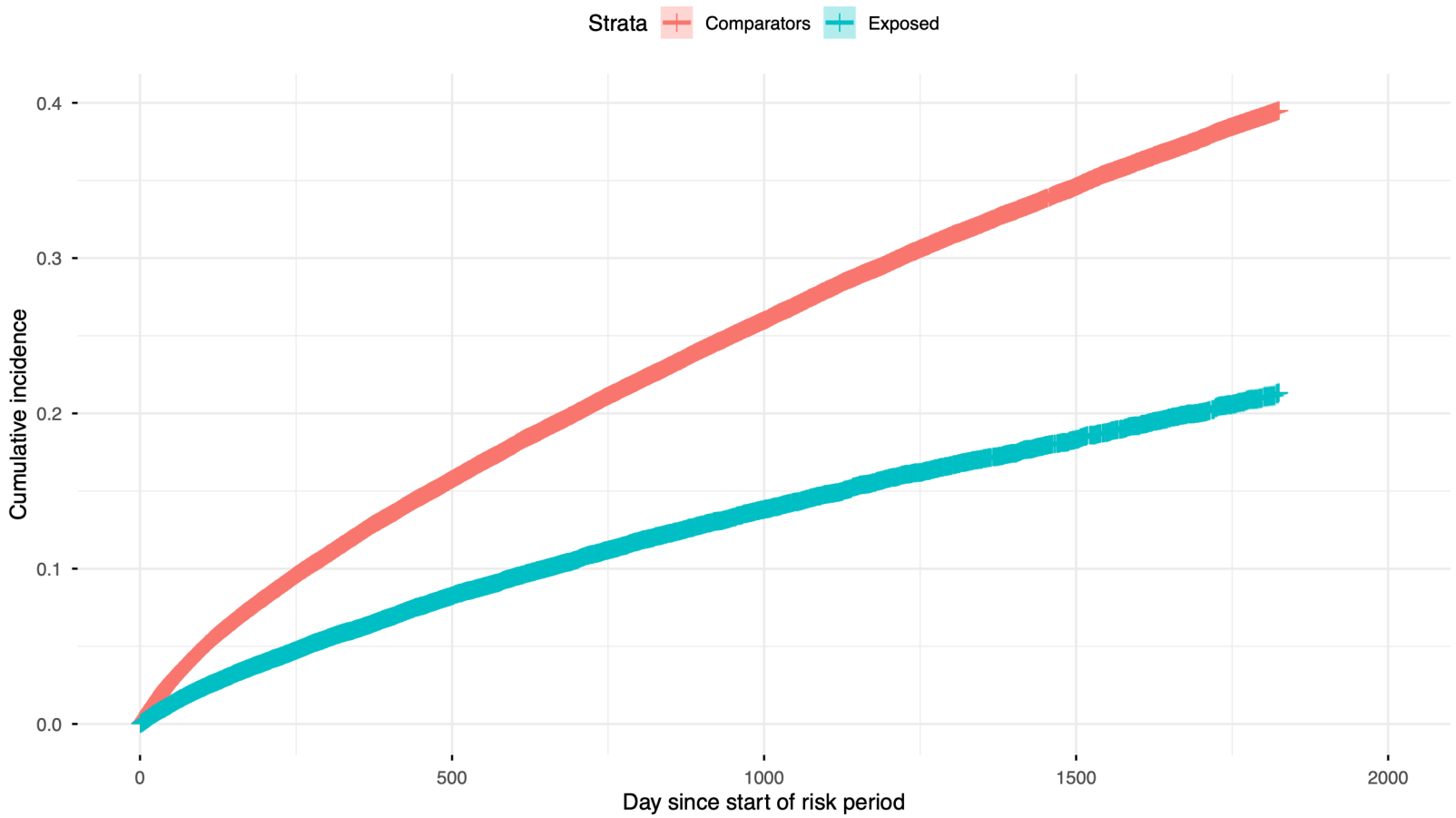
Comparison	Dapagliflozin						DPP4-I					
	N patients	N events	Person-years	Incidence rate (95% CI)	RMST year 3 (95% CI)	RMST year 5 (95% CI)	N patients	N events	Person-years	Incidence rate (95% CI)	RMST year 3 (95% CI)	RMST year 5 (95% CI)
Estimand 1	71159	6349	111223.08	57.08 (55.7–58.5)	N/A	N/A	134834	33239	301266.64	110.33 (109.1–111.5)	N/A	N/A
Estimand 2	71159	4518	79924.95	56.53 (54.9–58.2)	N/A	N/A	134834	19886	184300.56	107.9 (106.4–109.4)	N/A	N/A
Estimand 3**	71159	5846	98609.61	N/A	2.75 (2.74–2.75)	4.38 (4.37–4.39)	134834	28225	244690.4	115.35 (114–116.7)	2.52 (2.51–2.52)	3.84 (3.83–3.85)

*Person years: total follow-up time accumulated across all patients
Note: For Estimands 1 and 2, the number of events and accumulated person-time are reported over the maximum 5-year follow-up, whereas for Estimand 3 these quantities are reported over 3 years of follow-up only.



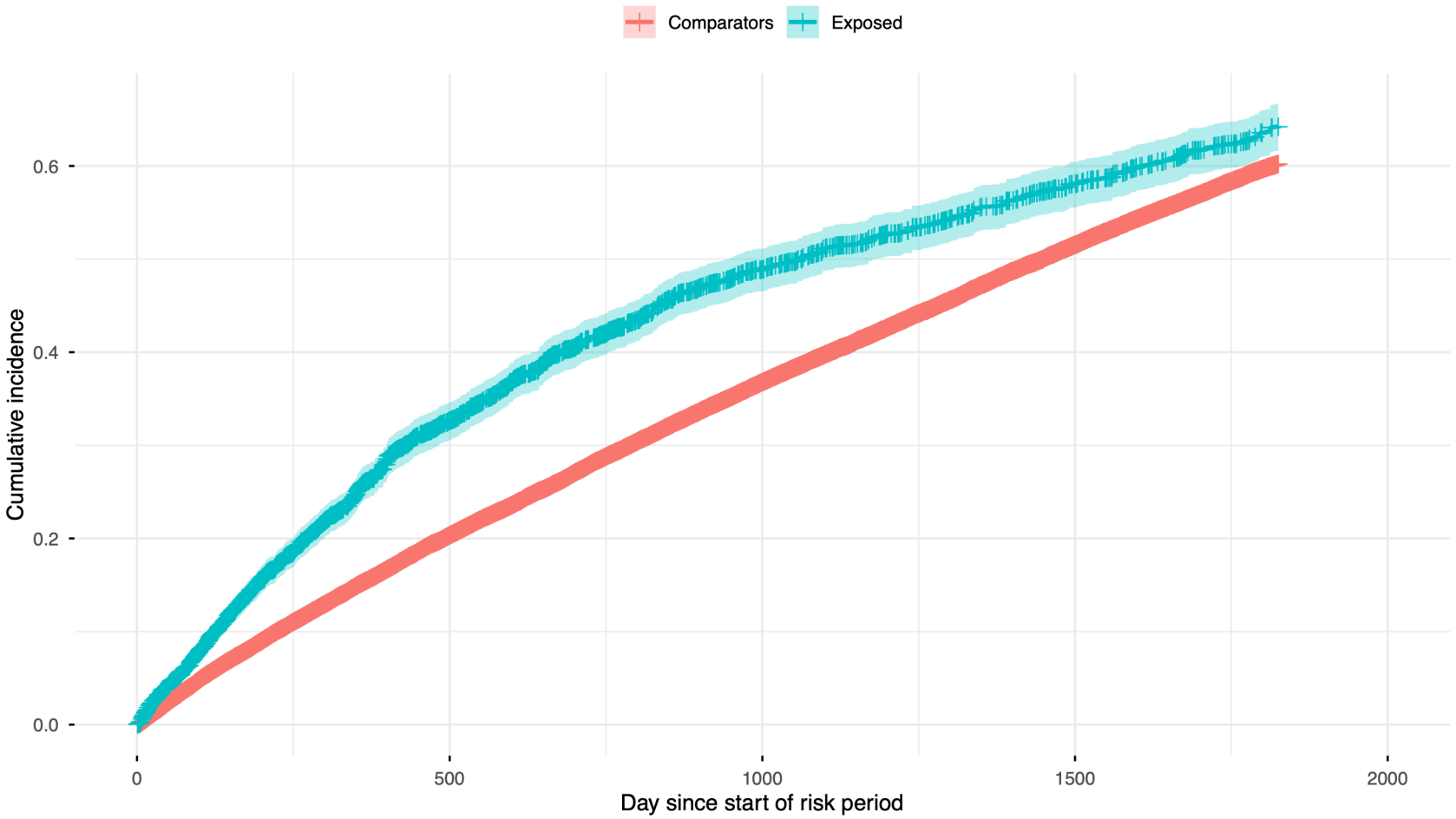
*Exposed is dapagliflozin, comparator is Dpp4-i

Figure 5. Crude cumulative incidence curves for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors) under Estimand 1, CPRD



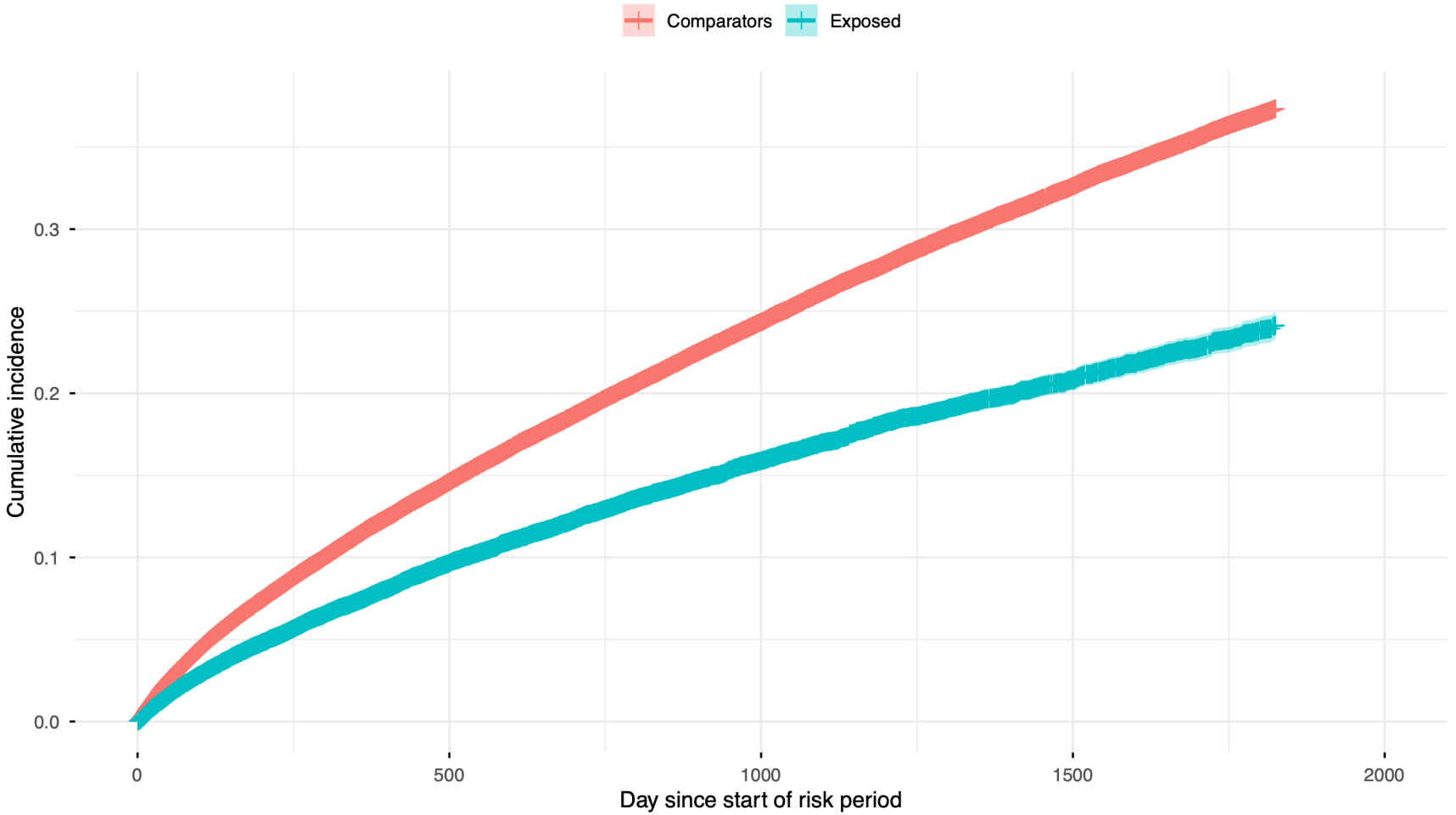
*Exposed is dapagliflozin, Comparator is Dpp4-i

Figure 6. Crude cumulative incidence curves for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors) under Estimand 1, BIFAP



*Exposed is dapagliflozin, comparator is Dpp4-i

Figure 7. Adjusted cumulative incidence curves for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors) under Estimand 1, CPRD



*Exposed is dapagliflozin, comparator is Dpp4-i

Figure 8. Adjusted cumulative incidence curves for MACE by treatment group (dapagliflozin vs DPP-4 inhibitors) under Estimand 1, BIFAP

10.4. *Primary Estimand*

In CPRD (Table 25), dapagliflozin was associated with a higher hazard of MACE than DPP4-i in both the crude (HR 1.41, 95% CI 1.34–1.49) and IPTW-adjusted (HR 1.30 95% CI 1.20–1.40) models. A modest attenuation toward the null after adjustment for measured baseline imbalances was observed, but no reversal of direction. This pattern is visible in the IPTW-adjusted cumulative incidence curves (

*Exposed is dapagliflozin, comparator is Dpp4-i

Figure 7): the exposed (dapagliflozin) curve separates above the comparator (DPP4-i) curve early in follow-up and remains higher throughout the risk period, with the gap widening over the first year before the two curves tend to converge from day 1000 until the end of the follow-up.

In BIFAP (Table 26), dapagliflozin was associated with a lower hazard of MACE than DPP4-i in both the crude (HR 0.49, 95% CI 0.48–0.50) and IPTW-adjusted (HR 0.62, 95% CI 0.61–0.64) models. This pattern is visible in the IPTW-adjusted cumulative incidence curves (*Exposed is dapagliflozin, comparator is Dpp4-i

Figure 8): the comparator (DPP4-i) curve separates above the exposed (dapagliflozin) curve early in follow-up and remains higher throughout the risk period, with the gap widening steadily across the full follow-up.

10.4.1. *Primary analyses*

Table 25. Crude and adjusted hazard ratios (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 1), CPRD

Comparison	Estimand	Crude HR (95% CI)	Adjusted HR (95% CI)
Dapagliflozin vs DPP4-I	Estimand 1	1.41 (1.34–1.49)	1.30 (1.20–1.40)

Table 26. Crude and adjusted hazard ratios (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 1), BIFAP

Comparison	Estimand	Crude HR (95% CI)	Adjusted HR (95% CI)
Dapagliflozin vs DPP4-I	Estimand 1	0.49 (0.48–0.50)	0.62 (0.61–0.64)

10.4.2. *Sensitivity analyses*

10.4.2.1. *Sensitivity analyses to missing data assumptions*

In CPRD, the adjusted HR for MACE (dapagliflozin vs DPP4-i) was materially unchanged when missing HbA1c was handled via multiple imputation (HR 1.30, 95% CI 1.21–1.38) compared with the primary IPTW-adjusted analysis (HR 1.30, 95% CI 1.20–1.40) (Table 27).

Table 27. Adjusted hazard ratios (95% CI) for MACE comparing dapagliflozin with DPP-4 inhibitors across sensitivity analyses related to missing data, CPRD

Comparison	Analysis	Database	Adjusted HR (95% CI)
Dapagliflozin vs DPP4-I	Primary (estimand 1)	CPRD	1.30 (1.20–1.40)*
Dapagliflozin vs DPP4-I	Multiple Imputation for HbA1c	CPRD	1.29 (1.21-1.38)

*IPTW adjusted from primary analysis

10.4.2.2. Sensitivity analyses to other modelling assumptions

Correcting for non-differential exposure misclassification (

Table 28) pushes both estimates further from the null in their original direction: CPRD rises from 1.30 (1.20–1.40) to 1.56 (1.36–1.94); BIFAP falls from 0.62 (0.61–0.64) to 0.54 (0.49–0.59).

Table 28. Adjusted hazard ratios (95% CI) for MACE comparing dapagliflozin with DPP-4 inhibitors across sensitivity analyses related to other modelling assumptions, CPRD, BIFAP

Comparison	Analysis	Database	Adjusted HR (95% CI/SI)#
Dapagliflozin vs DPP4-I	Primary (estimand 1)	CPRD	1.30 (1.20–1.40)*

Comparison	Analysis	Database	Adjusted HR (95% CI/SI)#
Dapagliflozin vs DPP4-I	Non-Differential exposure misclassification correction applied to estimand 1	CPRD	1.56 (1.36- 1.94)**
Dapagliflozin vs DPP4-I	Primary (estimand 1)	BIFAP	0.62 (0.61–0.64)*
Dapagliflozin vs DPP4-I	Non-Differential exposure misclassification correction applied to estimand 1	BIFAP	0.54 (0.49-0.59)**

*IPTW adjusted from Primary analysis

** IPTW Adjusted HR from primary analysis corrected for non-differential exposure misclassification

For exposure misclassification analyses, the SI is a simulation intervals (2.5th and 97.5th percentiles of the Monte Carlo draws)

10.5. *Supplemental Estimand(s)*

Under estimand 2, in CPRD, dapagliflozin was associated with a higher hazard of MACE than DPP4-i (crude HR 1.65, 95% CI 1.55–1.77; adjusted HR 1.51, 95% CI 1.37–1.66), a larger effect than under Estimand 1. In BIFAP, dapagliflozin remained protective (crude HR 0.50, 95% CI 0.48–0.51; adjusted HR 0.63, 95% CI 0.61–0.66), essentially unchanged from Estimand 1.

Estimand 3 RMST differences, calculated at 3 and 5 years, were directionally consistent with the hazard ratio findings (Table 31 & Table 32). In CPRD, dapagliflozin was associated with less MACE-free survival time than DPP4-i at both horizons (crude difference -0.29, 95% CI -0.33 to -0.25 at 3 years; -0.45, 95% CI -0.53 to -0.38 at 5 years), attenuated modestly after adjustment (-0.23 and -0.34, respectively) but not reversed. In BIFAP, the direction was opposite: dapagliflozin was associated with more MACE-free survival time than DPP4-i (crude difference +0.23, 95% CI 0.22–0.24 at 3 years; +0.54, 95% CI 0.53–0.56 at 5 years), again attenuated after adjustment (+0.15 and +0.37, respectively) but not reversed. In both databases, the magnitude of the RMST difference increased with longer follow-up (3 to 5 years), and all findings are directionally consistent with the corresponding hazard ratios and crude incidence rate ratios reported above.

Table 29. Crude and adjusted hazard ratios (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 2), CPRD

Comparison	Estimand	Crude HR (95% CI)	Adjusted HR (95% CI)
Dapagliflozin vs DPP4-I	Estimand 2	1.65 (1.59–1.77)	1.51 (1.37–1.66)

Table 30. Crude and adjusted hazard ratios (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 2), BIFAP

Comparison	Estimand	Crude HR (95% CI)	Adjusted HR (95% CI)
Dapagliflozin vs DPP4-I	Estimand 2	0.50 (0.48–0.51)	0.63 (0.61–0.66)

Table 31. Crude and adjusted differences in restricted mean survival time (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 3), CPRD

Comparison	Crude Difference in RMST 3 year (95% CI)	Adjusted Difference in RMST 3 year (95% CI)	Crude Difference in RMST 5 year (95% CI)	Adjusted Difference in RMST 5 year (95% CI)
Dapagliflozin vs DPP4-I	-0.29 (-0.33 to -0.25)	-0.23 (-0.28 to -0.18)	-0.45 (-0.53 to -0.38)	-0.34 (-0.44 to -0.22)

Table 32. Crude and adjusted differences in restricted mean survival time (95% CI) for the effect of dapagliflozin vs DPP-4 inhibitors on MACE (Estimand 3), BIFAP

Comparison	Crude Difference in RMST 3 year (95% CI)	Adjusted Difference in RMST 3 year (95% CI)	Crude Difference in RMST 5 year (95% CI)	Adjusted Difference in RMST 5 year (95% CI)
Dapagliflozin vs DPP4-I	0.23 (0.22–0.24)	0.15 (0.14–0.16)	0.54 (0.53–0.56)	0.37 (0.35–0.39)

11. Additional Analyses

None

12. Discussion

12.1. *Key Results*

The primary estimand targeted the following research question: what is the hazard ratio of MACE for dapagliflozin versus DPP4-i in patients with type 2 diabetes with, or at risk for, ASCVD, regardless of treatment discontinuation, switching, or new add-on antihyperglycaemic therapy? The direction of the crude and IPTW-adjusted association differed between databases and was consistent across estimands and analytic approaches. Under the primary

estimand, dapagliflozin was associated with a higher hazard of MACE than DPP4-i in CPRD (adjusted HR 1.30, 95% CI 1.20–1.40) but a lower hazard in BIFAP (adjusted HR 0.62, 95% CI 0.61–0.64). This reversal was not the result of the primary estimand's treatment-policy definition: it was identical in Estimand 2 (adjusted HR 1.51 in CPRD vs. 0. 0.63 in BIFAP), the crude incidence rates, the cumulative incidence curves, and the RMST differences (Estimand 3). Absolute incidence rates were also substantially higher in CPRD than in BIFAP in both treatment arms, and the composition of the composite outcome differed between databases, with death contributing a much larger share of events in CPRD than in BIFAP.

12.2. *Interpretation of results*

This case study clearly illustrates how the estimand framework clarifies intercurrent events handling from the estimand, to the design, the analysis and to the reporting in the context of target trial emulation. At the estimand stage, Estimand 1 treats discontinuation, switching, and add-on therapy the same way (treatment policy), applying it as a deliberate, justified choice rather than simply the default option it can represent, while applying a different intercurrent-event strategy to all-cause death: the composite strategy, which incorporates death directly into the outcome definition. This is a good real-world response to critique that the estimand framework conflates exposure and outcome — the concern being that treatment-related events (like stopping a medication) and outcome-related events (like death) get grouped into the same IE structure even though they are conceptually different things. Here they are not grouped together: the treatment-related events use a treatment policy strategy, while death is handled through the composite strategy instead. At the design stage, the estimand's choices shape concrete decisions: discontinuation is defined using a 90-day gap in prescription refills and the emulation tables tie each deviation between the target trial and the emulation to a specific part of the estimand, such as tobacco under-reporting affecting the population definition or the refill-gap assumption affecting how IEs are identified.

A second emulation table extends this same idea to the analysis stage itself, documenting the deviations between the analysis that would be conducted in trial vs what we can emulate using real-world data (e.g., weighting as an imperfect substitute for randomization), tying each deviation back to the specific assumption it addresses and keeping it visible so that it is documented explicitly rather than left implicit. Additionally, at the analysis stage, the intercurrent events in the analysis core table show the same events treated differently depending on the estimand: database end, end of study, and loss to follow-up are treated as censoring under every estimand, since these reflect features of the data and study design rather than the estimand's intercurrent-event strategy. By contrast, how discontinuation, switching, and add-on therapy are handled is estimand-specific: they are ignored under Estimand 1's treatment-policy strategy, but treated as censoring events under Estimand 2's while-on-treatment strategy, since follow-up after the event is no longer relevant to that estimand rather than missing. Because this censoring structure is defined in advance for each estimand, the analysis plan could also target its sensitivity checks precisely. At the reporting stage, this is made visible by reporting event counts and time-to-event separately for each event type and each estimand, instead of lumping all censoring together.

Confounding by indication is not fully explained by measured baseline differences. Pre-weighting, dapagliflozin initiators were younger, more often male, and more frequently prescribed insulin and aldosterone antagonists than DPP4-i initiators, particularly in CPRD; IPTW achieved near-complete

balance on these measured covariates (ASD ≤0.05 in BIFAP). This only addresses measured covariates, however, and does not rule out an unmeasured severity marker such as heart failure status itself — distinct from the proxy medications captured in the baseline tables — which the evidence below supports.

A further contributor, and one that helps explain why the reversal is more pronounced in CPRD, is the marked difference in absolute mortality between databases. MACE incidence was substantially higher in CPRD than BIFAP in both arms, and death accounted for 84–93% of CPRD MACE events versus 55–84% in BIFAP. Because all three estimands include all-cause death directly into the composite rather than modelling it as a competing risk, the treatment comparison is highly sensitive to each database's overall mortality level, irrespective of cause.

This composite strategy assumes all-cause death is a reasonable proxy for cardiovascular death in an ASCVD population, an assumption that may not have been reasonable given the COVID-19 pandemic, which increased all-cause mortality for largely non-cardiovascular reasons. If a meaningful share of CPRD's composite events were COVID-19 deaths unrelated to cardiovascular disease or drug exposure, this could inflate CPRD event rates in both arms and, if COVID-19 mortality risk differed systematically between arms — for example via frailty, comorbidity burden, or calendar timing of cohort entry — bias the treatment contrast itself.

Taken together, we consider it likely that the observed cross-database reversal reflects real differences in the population of dapagliflozin initiators over the study period. As dapagliflozin's indications expanded to include high cardiovascular risk and, from November 2020, heart failure, later CPRD initiators plausibly included a growing share of more advanced heart failure patients channeled to dapagliflozin as first-line therapy. Because this shift overlapped substantially with the COVID-19 pandemic, and heart failure patients carry elevated COVID-19 mortality risk, this channeling could explain a real mortality increase among dapagliflozin initiators in CPRD — captured in the composite through the death-handling strategy — rather than reflecting an isolated analytic artifact or another source of bias.

This pattern likely reflects the evolution of dapagliflozin's licensed indications: approved for type 2 diabetes in 2012, its label was updated in August 2019 for cardiovascular risk reduction (DECLARE-TIMI 58), it gained its first heart failure indication (HFrEF, with or without diabetes) in November 2020, and the MHRA extended this in December 2022 to the full LVEF spectrum. These successive expansions are consistent with the channeling pattern described above.

This channeling is supported by CPRD's baseline characteristics: aldosterone-antagonist use (a proxy for established HFrEF) is nearly three times higher among dapagliflozin than DPP-4i initiators before weighting (19.4% vs 6.4%), alongside elevated beta-blocker, diuretic, and ischaemic heart disease prevalence. This imbalance is far weaker in BIFAP, where the aldosterone-antagonist gap is roughly a third the size and ischaemic heart disease is balanced — consistent with more pronounced channeling in CPRD specifically. Combined with the more-male, more-ischaemic, higher-MRA profile, this looks less like generically frailer patients and more like a specific HFrEF phenotype (younger, male, ischaemic aetiology) — clean fit with dapagliflozin's first HF indication being HFrEF specifically, before the 2022 extension broadened it to the older, more female, preserved-ejection-fraction population.

In an unpublished analysis of the same CPRD data cut with a slightly different estimand, we compared SGLT2 inhibitors with DPP-4 inhibitors on stroke and MI separately, handling death via a while-alive approach; the population and COVID-period overlap were identical, though propensity-score matching was used in place of IPTW. Results were in line with expectations and all other findings: stroke HR 0.87 (95% CI 0.774–0.989); MI HR 0.97 (95% CI 0.841–1.11).

This is informative because it holds the population and COVID-overlapping period constant while changing how death is handled — as a competing risk via while-alive, rather than included into a composite — and examining stroke and MI individually rather than combined with death. Both land where the broader SGLT2i evidence would predict, suggesting the excess CPRD risk is concentrated in how mortality enters the composite rather than reflecting generalised confounding from channeling: if channeling were driving broad confounding, cause-specific vascular outcomes would likely be distorted too, and they are not. This supports the explanation above — that the reversal reflects real excess mortality among a more advanced, HFrEF-channeled CPRD population being captured through the composite strategy, not a broader source of bias.

Prescribing trends support this differential-uptake explanation. Open prescribing data show a marked, sustained rise in dapagliflozin prescribing over the study period, consistent with its expanding indications, while the DPP-4i comparators have been comparatively stable — sitagliptin broadly flat with a recent uptick, linagliptin similarly stable. Later-period initiators therefore make up a disproportionate share of the dapagliflozin arm, reinforcing the calendar-period and channeling arguments above.

This interpretation also fits the broader evidence base. Randomised clinical trial evidence for dapagliflozin's effect on MACE is neutral (DECLARE-TIMI 58: 8.8% vs 9.4%; HR 0.93, 95% CI 0.84–1.03) [1], while observational comparisons with DPP-4 inhibitors consistently favour SGLT2 inhibitors, in BIFAP's direction but at smaller magnitude: CVD-REAL Nordic HR 0.79 for MACE [12], CVD-REAL 2 (13 countries, including Spain) HR 0.88 for MI and 0.85 for stroke [13], and a US target trial emulation HR 0.88 for MACE [14] — none in a CPRD-like harmful direction.

Most relevant, a separate CPRD cohort study (Idris et al., 2021) [15] found SGLT2 inhibitors protective against mortality, heart failure, and CKD compared with DPP-4 inhibitors in the same UK data infrastructure. It did not use a MACE composite, so the comparison is qualitative, but its event rates are telling: in its ASCVD/high-risk subgroup (n=11,367/arm, 2013–2018), MI+stroke+death rates were ~44 (SGLT2i) vs ~50 (DPP4i) per 1,000 person-years — three- to five-fold below our CPRD Estimand 1 rates (237.2 vs 169.8) — and its study period predates COVID-19 entirely, consistent with our pandemic-mortality hypothesis.

Not all evidence favours SGLT2 inhibitors: a 42-RCT meta-analysis found no significant MACE difference versus DPP-4 inhibitors [16], and a 23-trial network meta-analysis found DPP-4 inhibitors, unlike SGLT2 inhibitors, did not reduce cardiorenal risk versus placebo [17]. Overall, though, the evidence points to a null-to-modest protective true effect, making our CPRD estimate an outlier — consistent with a database- or outcome-definition-specific artefact rather than genuine harm. Cause specific outcome analyses or removing covid-19 time period may address these uncertainties.

These results also bears on the choice of sensitivity analyses specified in the protocol. The pre-specified analysis addressing exposure misclassification targets misclassification of treatment status at index date only; correcting for it moved the estimates further from the null without

resolving the cross-database reversal. The sources of bias more consistent with the findings above — unobserved confounding, most plausibly an unmeasured heart failure severity marker, and differential outcome (mis)classification arising from how all-cause death enters the composite — were not directly targeted by any pre-specified sensitivity analysis. In retrospect, exposure misclassification, and by extension non-compliance with prescribing, appears to be the least important of the candidate sources of bias in this study. This mismatch reflects that the protocol and statistical analysis plan were finalised before data access, before the database-specific issues that emerged during analysis and reporting (particularly the interaction between the composite outcome's death-handling strategy and each database's mortality profile) could be foreseen. It points to two changes for future case studies of this design: allowing sensitivity analyses to be added post-hoc as issues are identified during analysis, rather than fixing them entirely at the protocol stage, and involving more data-source-specific expertise when the protocol is developed.

12.3. *Limitations*

Several elements of the target trial could not be fully replicated in its emulation and represent structural limitations of the design. Most importantly, the target trial's simple 1:1 randomisation — which would ensure balance in both measured and unmeasured confounders — could not be emulated; treatment assignment in this study instead reflects clinical need, with IPTW used to balance measured baseline confounders only, so unmeasured confounding cannot be excluded, and its likely direction (toward higher cardiovascular/renal risk in the dapagliflozin arm, particularly in CPRD) is consistent with the residual imbalance already discussed. Frailty was removed, MI was added as additional sensitivity analysis.

Eligibility criteria that rely on clinical judgement in the target trial, such as established or high ASCVD risk incorporating tobacco use, had to be operationalised using diagnostic and prescription codes in routinely collected data, introducing potential misclassification, most notably under-ascertainment of tobacco use. Follow-up in the emulation begins at first prescription rather than at randomisation, and loss to follow-up is proxied by deregistration, practice withdrawal, or database end rather than directly observed loss of contact.

Emulation of exposure and covariate data introduces further departures from the target trial. Treatment discontinuation is inferred from a >90-day gap in prescription refills, and missing exposure data are interpreted as true discontinuation after this window — an assumption not required in a trial with direct exposure monitoring, and one whose accuracy depends on the chosen gap threshold and lookback period. Missing covariate data are handled by assuming an absent diagnosis code reflects a true absence of the condition, and missing lifestyle or laboratory values are imputed under a missing-at-random assumption — mechanisms of missingness with no direct analogue in the target trial.

The actual event counts for Estimand 1 (1,731 dapagliflozin/14,403 DPP4-I events in CPRD; 6,349/33,239 in BIFAP) far exceeded the 823–1,189 events assumed in the precision-based sample size scenarios, but the original calculation assumed 1:1 allocation, whereas actual treatment allocation was far from equal — roughly 1:8 in CPRD and 1:2 in BIFAP, a direct consequence of real-world prescribing. Redoing the precision estimate to account for this actual imbalance explains why: precision comes out worse than a simple calculation based on total events alone would suggest,

because CPRD has so few dapagliflozin users that they limit how much information the analysis actually has, even with a large DPP4-I arm. CPRD's actually observed adjusted-model precision (≈7.7%) was wider still, suggesting a further efficiency loss from covariate adjustment given how thin the exposed arm is, while BIFAP's observed precision (≈2.9%) tracked its unequal-allocation prediction closely. In practice, then, both databases ended up at least as precise as planned, and BIFAP substantially more so.

13. Conclusion

In this two-database study, applying a formal estimand framework — treatment policy (Estimand 1), while-on-treatment (Estimand 2), and RMST (Estimand 3) — showed that the association between dapagliflozin and MACE relative to DPP4 inhibitors differed in direction between CPRD and BIFAP consistently across all three estimands, and was not resolved by adjustment for measured confounders. Because the reversal persisted regardless of how intercurrent events such as discontinuation and switching were handled, the estimand framework let us rule out the analytic definition of the estimand itself as an explanation. This points instead to setting-specific differences in patient populations, prescribing practice, or outcome occurrence (most plausibly a more advanced, heart-failure-channelled dapagliflozin population in CPRD interacting with the composite outcome's death-handling strategy during the COVID-19 period) rather than a consistent causal effect of dapagliflozin on cardiovascular risk relative to DPP4-i.

More broadly, this case study demonstrates the practical value of a pre-specified estimand framework in real-world evidence generation. By making intercurrent-event strategies explicit at every stage, from estimand definition through design, analysis, and reporting, it distinguishes genuine cross-database effect heterogeneity from an artefact of analytic choice, and keeps assumptions that would otherwise stay implicit more explicit.

14. 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;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. doi: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;111(1):321–331.

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

12. Persson F, Nyström T, Jørgensen ME, et al. Dapagliflozin is associated with lower risk of cardiovascular events and all-cause mortality in people with type 2 diabetes (CVD-REAL Nordic) when compared with dipeptidyl peptidase-4 inhibitor therapy: A multinational observational study. Diabetes Obes Metab. 2018;20(2):344–351. doi:10.1111/dom.13077

13. Kosiborod M, et al; CVD-REAL 2 Investigators and Study Group. Risk of cardiovascular events and death associated with initiation of SGLT2 inhibitors compared with DPP-4 inhibitors: an analysis from the CVD-REAL 2 multinational cohort study. Lancet Diabetes Endocrinol. 2020;8(7):606–615. doi:10.1016/S2213-8587(20)30130-3

14. Xie Y, Bowe B, Xian H, Loux T, McGill JB, Al-Aly Z. Comparative effectiveness of SGLT2 inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonylureas on risk of major adverse cardiovascular events: emulation of a randomised target trial using electronic health records. Lancet Diabetes Endocrinol. 2023;11(9):644–656. doi:10.1016/S2213-8587(23)00171-7

15. Idris I, Zhang R, Mamza JB, Ford M, Morris T, Banerjee A, Khunti K. Lower risk of hospitalization for heart failure, kidney disease and death with sodium-glucose co-transporter-2 inhibitors compared with dipeptidyl peptidase-4 inhibitors in type 2 diabetes regardless of prior cardiovascular or kidney disease: A retrospective cohort study in UK primary care. Diabetes Obes Metab. 2021;23(10):2207–2214. doi:10.1111/dom.14437

16. Li Y, Chen J, Gao Y, et al. Comparative Safety of SGLT2 Versus DPP4 Inhibitors in Patients with Type 2 Diabetes: A Meta-Analysis of Randomized Controlled Trials. Diabetes Ther. 2026;17:685–709. doi:10.1007/s13300-025-01836-2

17. Giugliano D, Longo M, Signoriello S, Maiorino MI, Solerte B, Chiodini P, Esposito K. The effect of DPP-4 inhibitors, GLP-1 receptor agonists and SGLT-2 inhibitors on cardiorenal outcomes: a network meta-analysis of 23 CVOTs. Cardiovasc Diabetol. 2022;21:42. doi:10.1186/s12933-022-01474-z

15. Appendices

15.1. *Supplemental Tables and Figures*

Supplemental Table 1. Attrition Table, CPRD

Attrition step	N	%
All individuals registered in CPRD without missing sex and have valid year of birth during study period	**517,173**	**100**
AND received SGLT2-I or Dpp4-I during enrolment period (2013-2024)	77,707	15.03
AND have linkage with hospital data	77,707	15.03
AND received dapagliflozin as SGLT2-i	72,033	13.93
AND have at least one year of recorded medical history in database prior to dapagliflozin or DPP4i	72,033	13.93
AND DO NOT have a documented history or diagnosis of Type 1 diabetes at any time prior to index	56,288	10.88
AND are aged 40 or older at index	56,288	10.88
AND have a diagnosis of T2D prior to index	51,185	9.90
AND have either established ACVD or high ACVD risk	43,644	8.44
AND WERE NOT treated with an SGLT2-I or DPP4-I in previous year	36,288	7.02
AND DO NOT have a documented history of acute cardiovascular event in the year prior to index	34,110	6.60
Total Eligible Cohort	**34,110**	
Eligible dapagliflozin	**3,850**	
Eligible DPP4i	**30,260**	

Supplemental Table 2. Attrition Table, BIFAP

Attrition step	N	%
All individuals registered in the data source	**1,019,538**	**—**
AND have at least one episode in primary care	1,017,038	99.75

Attrition step	N	%
All individuals registered in the primary care data source with complete sex, death date and at least one episode in primary care	**996,329**	**97.72**
AND have a valid year of birth	996,328	97.72
AND have linkage with hospital data	996,328	97.72
AND are registered during the study period (2012 to 2023)	996,328	97.72
AND received dapagliflozin or DPP4i during the enrolment period 1 Jan 2013 to 31 December 2023	795,035	77.98
AND have at least one year of recorded medical history in database prior to dapagliflozin or DPP4i	571,657	56.07
AND DO NOT have a documented history or diagnosis of Type 1 diabetes at any time prior to index	559,296	54.86
AND are aged 40 or older at index	552,270	54.17
AND have a diagnosis of T2D prior to index	327,237	32.10
AND have either established ACVD or high ACVD risk	279,074	27.37
AND WERE NOT treated with an SGLT2-I or DPP4-I in previous year	219,003	21.48
AND DO NOT have a documented history of acute cardiovascular event in the year prior to index	205,993	20.20
Total Eligible Cohort	**205,993**	
Eligible dapagliflozin	**71,159**	
Eligible DPP4i	**134,834**	

Supplemental Table 3. Baseline patient characteristics of the study population before imputation, CPRD

Characteristic	Overall	Dapagliflozin	DPP4-i
N	**34110**	**3850**	**30260**
Characteristics			
Age (years)			
Mean (SD)	74.6 (10.1)	70.6 (10.5)	75.1 (9.9)
Median (IQ Range)	75 (67-82)	69 (62-78)	76 (68-83)
Age (years) N (%)			
<65	6262 (18.4%)	1266 (32.9%)	4996 (16.5%)
65-74	9961 (29.2%)	1234 (32.1%)	8727 (28.8%)

Characteristic	Overall	Dapagliflozin	DPP4-i
75-84	11900 (34.9%)	869 (22.6%)	11031 (36.5%)
≥85	5987 (17.6%)	481 (12.5%)	5506 (18.2%)
Gender N (%)			
Male	19977 (58.6%)	2637 (68.5%)	17340 (57.3%)
Female	14132 (41.4%)	1213 (31.5%)	12919 (42.7%)
Other	< 5 (%)	0 (0%)	< 5 (%)
Calendar year at t0			
2013	4710 (13.8%)	104 (2.7%)	4606 (15.2%)
2014	4458 (13.1%)	359 (9.3%)	4099 (13.5%)
2015	4656 (13.6%)	361 (9.4%)	4295 (14.2%)
2016	4426 (13%)	240 (6.2%)	4186 (13.8%)
2017	3914 (11.5%)	195 (5.1%)	3719 (12.3%)
2018	3214 (9.4%)	154 (4%)	3060 (10.1%)
2019	2530 (7.4%)	163 (4.2%)	2367 (7.8%)
2020	1665 (4.9%)	139 (3.6%)	1526 (5%)
2021	1646 (4.8%)	436 (11.3%)	1210 (4%)
2022	1487 (4.4%)	783 (20.3%)	704 (2.3%)
2023	1074 (3.1%)	698 (18.1%)	376 (1.2%)
2024	330 (1%)	218 (5.7%)	112 (0.4%)
2025	0 (0%)	0 (0%)	0 (0%)
Duration of diabetes			
Mean (SD)	8.8 (5.3)	8.4 (5.7)	8.9 (5.3)
Median (IQ range)	8.6 (4.6-12.7)	7.8 (3.7-12.3)	8.7 (4.7-12.7)
Missing data pre-imputation			
HbA1c			
Mean (SD)	70.3 (18.6)	68.1 (21.2)	70.5 (18.2)
Median (IQ range)	67 (58-79)	64 (51.8-81)	67 (59-79)
Missing N (%)	933 (2.7%)	154 (4%)	779 (2.6%)

Characteristic	Overall	Dapagliflozin	DPP4-i
eGFR			
Mean (SD)	67.8 (24.3)	76.6 (22.6)	66.7 (24.3)
Median (IQ range)	69.8 (47.3-89.4)	81.1 (60.1-95.1)	68.1 (46.2-88.3)
Missing N (%)	731 (2.1%)	94 (2.4%)	637 (2.1%)
Tobacco use			
None	9978 (29.3%)	1027 (26.7%)	8951 (29.6%)
Past	20438 (59.9%)	2340 (60.8%)	18098 (59.8%)
Current	3593 (10.5%)	474 (12.3%)	3119 (10.3%)
Missing N (%)	101 (0.3%)	9 (0.2%)	92 (0.3%)
(Co)Morbidities			
Obesity	2253 (6.6%)	410 (10.6%)	1843 (6.1%)
Heart transplant	116 (0.3%)	29 (0.8%)	87 (0.3%)
Microvascular complications	263 (0.8%)	34 (0.9%)	229 (0.8%)
Severe hypoglycaemia	< 5 (%)	0 (0%)	< 5 (%)
Diabetic neuropathy	1497 (4.4%)	133 (3.5%)	1364 (4.5%)
Diabetic retinopathy	10203 (29.9%)	965 (25.1%)	9238 (30.5%)
Keto-/lactate acidosis	91 (0.3%)	5 (0.1%)	86 (0.3%)
Lower limb amputations	359 (1.1%)	33 (0.9%)	326 (1.1%)
COPD	5854 (17.2%)	725 (18.8%)	5129 (16.9%)
Cancer	8777 (25.7%)	933 (24.2%)	7844 (25.9%)
Major organ specific bleeding	643 (1.9%)	70 (1.8%)	573 (1.9%)
Bariatric surgery	30 (0.1%)	6 (0.2%)	24 (0.1%)
CKD stages 1-4	27763 (81.4%)	3108 (80.7%)	24655 (81.5%)
CKD stage 5	5403 (15.8%)	571 (14.8%)	4832 (16%)
Kidney transplantation	120 (0.4%)	< 5 (%)	116 (0.4%)
Dialysis	98 (0.3%)	< 5 (%)	95 (0.3%)
Comedications			
Metformin	26675 (78.2%)	2835 (73.6%)	23840 (78.8%)

Characteristic	Overall	Dapagliflozin	DPP4-i
Sulfonylurea	15430 (45.2%)	1131 (29.4%)	14299 (47.3%)
Sulfonamides	3124 (9.2%)	165 (4.3%)	2959 (9.8%)
Alfa glucosidase inhibitors	96 (0.3%)	6 (0.2%)	90 (0.3%)
Thiazolidinediones	1948 (5.7%)	187 (4.9%)	1761 (5.8%)
Other GLDs	140 (0.4%)	13 (0.3%)	127 (0.4%)
Insulin	2731 (8%)	548 (14.2%)	2183 (7.2%)
ACE inhibitors	16798 (49.2%)	2038 (52.9%)	14760 (48.8%)
ARBs	7348 (21.5%)	818 (21.2%)	6530 (21.6%)
Beta-blockers	13334 (39.1%)	1773 (46.1%)	11561 (38.2%)
Low/high ceiling diuretics	9757 (28.6%)	1367 (35.5%)	8390 (27.7%)
Aldosterone antagonists	2674 (7.8%)	746 (19.4%)	1928 (6.4%)
Thiazide diuretics	3891 (11.4%)	305 (7.9%)	3586 (11.9%)
Calcium channel blockers	13371 (39.2%)	1300 (33.8%)	12071 (39.9%)
Digitoxin/digoxin	2407 (7.1%)	268 (7%)	2139 (7.1%)
Antiarrhythmics	452 (1.3%)	94 (2.4%)	358 (1.2%)
Statins	27391 (80.3%)	3105 (80.6%)	24286 (80.3%)
Warfarin	3364 (9.9%)	228 (5.9%)	3136 (10.4%)
Aspirin	11865 (34.8%)	1171 (30.4%)	10694 (35.3%)
Purinergic receptor $P2Y_{12}$ antagonists	3082 (9%)	348 (9%)	2734 (9%)
Other antiplatelets	11869 (34.8%)	1170 (30.4%)	10699 (35.4%)
Corticosteroids	5714 (16.8%)	656 (17%)	5058 (16.7%)
Weight loss drugs	216 (0.6%)	41 (1.1%)	175 (0.6%)
Established ACVD			
Ischemic heart disease	4999 (14.7%)	719 (18.7%)	4280 (14.1%)
Ischemic cerebrovascular disease	629 (1.8%)	56 (1.5%)	573 (1.9%)
Peripheral arterial disease	339 (1%)	57 (1.5%)	282 (0.9%)
ACVD risk factors			
Dyslipidaemia	3172 (9.3%)	378 (9.8%)	2794 (9.2%)

Characteristic	Overall	Dapagliflozin	DPP4-i
Hypertension	14723 (43.2%)	1715 (44.5%)	13008 (43%)

Supplemental Table 4. Patient Characteristics of the study population before imputation, BIFAP

Characteristic	Overall	Dapagliflozin	DPP4-i
N	**205993**	**71159**	**134834**
Characteristics			
Age (years)			
Mean (SD)	74.6 (10.3)	71.7 (9.7)	76.1 (10.3)
Median (IQ Range)	75 (67-83)	71 (64-79)	77 (68-84)
Age (years) N (%)			
<65	39785 (19.3%)	18563 (26.1%)	21222 (15.7%)
65-74	61676 (29.9%)	24883 (35%)	36793 (27.3%)
75-84	63597 (30.9%)	19796 (27.8%)	43801 (32.5%)
≥85	40935 (19.9%)	7917 (11.1%)	33018 (24.5%)
Gender N (%)			
Male	108518 (52.7%)	42787 (60.1%)	65731 (48.7%)
Female	97475 (47.3%)	28372 (39.9%)	69103 (51.3%)
Other	0 (0%)	0 (0%)	0 (0%)
Calendar year at t0			
2013	1729 (0.8%)	0 (0%)	1729 (1.3%)
2014	2355 (1.1%)	212 (0.3%)	2143 (1.6%)
2015	4830 (2.3%)	653 (0.9%)	4177 (3.1%)
2016	7154 (3.5%)	830 (1.2%)	6324 (4.7%)
2017	9185 (4.5%)	1158 (1.6%)	8027 (6%)
2018	10643 (5.2%)	1440 (2%)	9203 (6.8%)
2019	16679 (8.1%)	2922 (4.1%)	13757 (10.2%)
2020	22198 (10.8%)	5065 (7.1%)	17133 (12.7%)

Characteristic	Overall	Dapagliflozin	DPP4-i
2021	34275 (16.6%)	11038 (15.5%)	23237 (17.2%)
2022	41701 (20.2%)	18747 (26.3%)	22954 (17%)
2023	55244 (26.8%)	29094 (40.9%)	26150 (19.4%)
2024	0 (0%)	0 (0%)	0 (0%)
2025	0 (0%)	0 (0%)	0 (0%)
Duration of diabetes			
Mean (SD)	4.1 (3.7)	4.7 (3.8)	3.8 (3.6)
Median (IQ range)	3.3 (0.8-6.6)	4 (1.3-7.3)	3 (0.6-6.2)
Missing data pre-imputation			
HbA1c			
Mean (SD)	57.7 (16.3)	59.3 (16.7)	56.9 (16)
Median (IQ range)	54.1 (46.5-65)	55.2 (47.5-67.2)	53 (46.5-63.9)
Missing N (%)	124988 (60.7%)	44586 (62.7%)	80402 (59.6%)
eGFR			
Mean (SD)	72 (23.4)	78.2 (20.6)	69 (24)
Median (IQ range)	75.6 (53.2-92)	82.7 (63.3-95.1)	71.9 (48.9-90.1)
Missing N (%)	122281 (59.4%)	44500 (62.5%)	77781 (57.7%)
Tobacco use			
None	253 (0.1%)	77 (0.1%)	176 (0.1%)
Past	22635 (11%)	8550 (12%)	14085 (10.4%)
Current	25547 (12.4%)	10730 (15.1%)	14817 (11%)
Missing N (%)	157558 (76.5%)	51802 (72.8%)	105756 (78.4%)
(Co)Morbidities			
Obesity	13080 (6.3%)	5046 (7.1%)	8034 (6%)
Heart transplant	163 (0.1%)	72 (0.1%)	91 (0.1%)
Microvascular complications	4060 (2%)	1695 (2.4%)	2365 (1.8%)
Severe hypoglycaemia	63 (0%)	32 (0%)	31 (0%)
Diabetic neuropathy	5373 (2.6%)	1968 (2.8%)	3405 (2.5%)

Characteristic	Overall	Dapagliflozin	DPP4-i
Diabetic retinopathy	8474 (4.1%)	3103 (4.4%)	5371 (4%)
Keto-/lactate acidosis	477 (0.2%)	121 (0.2%)	356 (0.3%)
Lower limb amputations	1215 (0.6%)	408 (0.6%)	807 (0.6%)
COPD	22600 (11%)	8104 (11.4%)	14496 (10.8%)
Cancer	36954 (17.9%)	11160 (15.7%)	25794 (19.1%)
Major organ specific bleeding	2602 (1.3%)	577 (0.8%)	2025 (1.5%)
Bariatric surgery	353 (0.2%)	184 (0.3%)	169 (0.1%)
CKD stages 1-4	90746 (44.1%)	28067 (39.4%)	62679 (46.5%)
CKD stage 5	16345 (7.9%)	3829 (5.4%)	12516 (9.3%)
Kidney transplantation	984 (0.5%)	219 (0.3%)	765 (0.6%)
Dialysis	655 (0.3%)	40 (0.1%)	615 (0.5%)
Comedications			
Metformin	92729 (45%)	27587 (38.8%)	65142 (48.3%)
Sulfonylurea	18357 (8.9%)	5703 (8%)	12654 (9.4%)
Sulfonamides	4730 (2.3%)	1077 (1.5%)	3653 (2.7%)
Alfa glucosidase inhibitors	598 (0.3%)	119 (0.2%)	479 (0.4%)
Thiazolidinediones	783 (0.4%)	312 (0.4%)	471 (0.3%)
Other GLDs	16894 (8.2%)	5400 (7.6%)	11494 (8.5%)
Insulin	36310 (17.6%)	12400 (17.4%)	23910 (17.7%)
ACE inhibitors	46379 (22.5%)	15858 (22.3%)	30521 (22.6%)
ARBs	42493 (20.6%)	14878 (20.9%)	27615 (20.5%)
Beta-blockers	66737 (32.4%)	25229 (35.5%)	41508 (30.8%)
Low/high ceiling diuretics	61318 (29.8%)	19614 (27.6%)	41704 (30.9%)
Aldosterone antagonists	19145 (9.3%)	8326 (11.7%)	10819 (8%)
Thiazide diuretics	6204 (3%)	1983 (2.8%)	4221 (3.1%)
Calcium channel blockers	48190 (23.4%)	15440 (21.7%)	32750 (24.3%)
Digitoxin/digoxin	7048 (3.4%)	2036 (2.9%)	5012 (3.7%)
Antiarrhythmics	5533 (2.7%)	2236 (3.1%)	3297 (2.4%)

Characteristic	Overall	Dapagliflozin	DPP4-i
Statins	125168 (60.8%)	44918 (63.1%)	80250 (59.5%)
Warfarin	910 (0.4%)	294 (0.4%)	616 (0.5%)
Aspirin	59089 (28.7%)	20588 (28.9%)	38501 (28.6%)
Purinergic receptor $P2Y_{12}$ antagonists	10779 (5.2%)	3621 (5.1%)	7158 (5.3%)
Other antiplatelets	60393 (29.3%)	20987 (29.5%)	39406 (29.2%)
Corticosteroids	30705 (14.9%)	9815 (13.8%)	20890 (15.5%)
Weight loss drugs	8 (0%)	< 5 (%)	5 (0%)
Established ACVD			
Ischemic heart disease	14114 (6.9%)	4969 (7%)	9145 (6.8%)
Ischemic cerebrovascular disease	7814 (3.8%)	1808 (2.5%)	6006 (4.5%)
Peripheral arterial disease	4052 (2%)	1170 (1.6%)	2882 (2.1%)
ACVD risk factors			
Dyslipidaemia	41808 (20.3%)	13363 (18.8%)	28445 (21.1%)
Hypertension	56237 (27.3%)	16454 (23.1%)	39783 (29.5%)

Supplemental Table 5. Description of the propensity score model parameters used to compute weights in the final model, CPRD

	Dpp4-i	Dapagliflozin	OR (95% CI)
N	**30,260**	**3,850**	
Demographics			
Age (years) N (%)			
< 65	4,996 (16.5%)	1,266 (32.9%)	Reference
65-74	8,727 (28.8%)	1,234 (32.1%)	0.532 (0.485–0.584)
75-84	11,031 (36.5%)	869 (22.6%)	0.26 (0.234–0.29)
85+	5,506 (18.2%)	481 (12.5%)	0.259 (0.226–0.296)
Gender N (%)			

	Dpp4-i	Dapagliflozin	OR (95% CI)
Male	17,340 (57.3%)	2,637 (68.5%)	1.433 (1.326–1.551)
Female	12,919 (42.7%)	1,213 (31.5%)	Reference
Other	< 5 (%)	0 (0%)	0 (–3653669.021)
Comorbidities N (%)			
Chronic Kidney Disease N (%)			
Stage 0	773 (2.6%)	171 (4.4%)	Reference
Stage 1-4	24,655 (81.5%)	3,108 (80.7%)	0.857 (0.716–1.031)
Stage 5	4,832 (16%)	571 (14.8%)	0.696 (0.566–0.86)
Obesity	1,843 (6.1%)	410 (10.6%)	1.352 (1.191–1.533)
Heart transplant	87 (0.3%)	29 (0.8%)	1.286 (0.797–2.02)
Microvascular complications	229 (0.8%)	34 (0.9%)	1.023 (0.671–1.513)
Severe hypoglycaemia	< 5 (%)	0 (0%)	0 (–8986351.795)
Diabetic neuropathy	1,364 (4.5%)	133 (3.5%)	0.766 (0.628–0.927)
Diabetic Retinopathy	9,238 (30.5%)	965 (25.1%)	0.815 (0.747–0.888)
Keto-/lactate acidosis	86 (0.3%)	5 (0.1%)	0.33 (0.113–0.764)
Lower limb amputations	326 (1.1%)	33 (0.9%)	0.692 (0.46–1.009)
COPD	5,129 (16.9%)	725 (18.8%)	1.048 (0.945–1.161)
Cancer	7,844 (25.9%)	933 (24.2%)	1.085 (0.996–1.181)
Major organ specific bleeding	573 (1.9%)	70 (1.8%)	0.705 (0.533–0.92)
Bariatric surgery	24 (0.1%)	6 (0.2%)	1.198 (0.411–2.997)
Kidney Transplantation	116 (0.4%)	< 5 (%)	0.217 (0.065–0.536)
Dialysis	95 (0.3%)	< 5 (%)	0.26 (0.062–0.726)
Comedications N (%)			
Metformin	23,840 (78.8%)	2,835 (73.6%)	0.62 (0.567–0.679)

	Dpp4-i	Dapagliflozin	OR (95% CI)
Sulfonylurea	14,299 (47.3%)	1,131 (29.4%)	0.499 (0.461–0.539)
Sulfonamides	2,959 (9.8%)	165 (4.3%)	0.508 (0.428–0.598)
Alfa glucosidase inhibitors	90 (0.3%)	6 (0.2%)	0.679 (0.255–1.494)
Thiazolidinediones	1,761 (5.8%)	187 (4.9%)	0.973 (0.824–1.143)
Other GLDs	127 (0.4%)	13 (0.3%)	0.77 (0.406–1.343)
Insulin	2,183 (7.2%)	548 (14.2%)	1.94 (1.73–2.174)
ACE inhibitors	14,760 (48.8%)	2,038 (52.9%)	1.196 (1.102–1.298)
ARBs	6,530 (21.6%)	818 (21.2%)	1.148 (1.04–1.267)
Beta-blockers	11,561 (38.2%)	1,773 (46.1%)	1.3 (1.198–1.411)
Low/high ceiling diuretics	8,390 (27.7%)	1,367 (35.5%)	1.307 (1.188–1.437)
Aldosterone antagonists	1,928 (6.4%)	746 (19.4%)	3.028 (2.699–3.397)
Thiazide diuretics	3,586 (11.9%)	305 (7.9%)	0.758 (0.665–0.861)
Calcium channel blockers	12,071 (39.9%)	1,300 (33.8%)	0.874 (0.809–0.944)
Digitoxin/digoxin	2,139 (7.1%)	268 (7%)	0.973 (0.83–1.137)
Antiarrhythmics	358 (1.2%)	94 (2.4%)	1.377 (1.052–1.788)
Statins	24,286 (80.3%)	3,105 (80.6%)	1.023 (0.932–1.124)
Warfarin	3,136 (10.4%)	228 (5.9%)	0.349 (0.297–0.409)
Aspirin	10,694 (35.3%)	1,171 (30.4%)	0.728 (0.397–1.33)
Purinergic receptor P2Y_12 antagonists	2,734 (9%)	348 (9%)	0.843 (0.741–0.957)
Other antiplatelets	10,699 (35.4%)	1,170 (30.4%)	0.927 (0.507–1.7)
Corticosteroids	5,058 (16.7%)	656 (17%)	0.994 (0.897–1.1)
Weight loss drugs	175 (0.6%)	41 (1.1%)	1.239 (0.849–1.766)
Lifestyle N (%)			
Tobacco use			

	Dpp4-i	Dapagliflozin	OR (95% CI)
None	9,043 (29.9%)	1,036 (26.9%)	Reference
Past	18,098 (59.8%)	2,340 (60.8%)	1.046 (0.962–1.138)
Current	3,119 (10.3%)	474 (12.3%)	0.946 (0.832–1.075)
Established ACVD			
Ischemic heart disease	4,280 (14.1%)	719 (18.7%)	1.237 (1.108–1.381)
Ischemic cerebrovascular disease	573 (1.9%)	56 (1.5%)	0.829 (0.611–1.106)
ACVD Risk factors			
Peripheral arterial disease	282 (0.9%)	57 (1.5%)	1.37 (0.978–1.892)
Dyslipidaemia	2,794 (9.2%)	378 (9.8%)	0.939 (0.826–1.065)
Hypertension	13,008 (43%)	1,715 (44.5%)	1.027 (0.948–1.113)

Supplemental Table 6. Description of the propensity score model parameters used to compute weights in the final model, BIFAP

	Dpp4-i	Dapagliflozin	OR (95% CI)
N	**134,834**	**71,159**	
Demographics			
Age (years) N (%)			
< 65	21,222 (15.7%)	18,563 (26.1%)	Reference
65-74	36,793 (27.3%)	24,883 (35%)	0.82 (0.798–0.842)
75-84	43,801 (32.5%)	19,796 (27.8%)	0.567 (0.551–0.583)
85+	33,018 (24.5%)	7,917 (11.1%)	0.319 (0.308–0.331)

	Dpp4-i	Dapagliflozin	OR (95% CI)
Gender N (%)			
Male	65,731 (48.7%)	42,787 (60.1%)	1.319 (1.292–1.347)
Female	69,103 (51.3%)	28,372 (39.9%)	Reference
Other	0 (0%)	0 (0%)	—
Comorbidities N (%)			
Chronic Kidney Disease N (%)			
Stage 0	59,639 (44.2%)	39,263 (55.2%)	Reference
Stage 1-4	62,679 (46.5%)	28,067 (39.4%)	0.747 (0.732–0.762)
Stage 5	12,516 (9.3%)	3,829 (5.4%)	0.565 (0.54–0.59)
Obesity	8,034 (6%)	5,046 (7.1%)	1.294 (1.242–1.347)
Heart transplant	91 (0.1%)	72 (0.1%)	1.214 (0.875–1.679)
Microvascular complications	2,365 (1.8%)	1,695 (2.4%)	1.235 (1.151–1.325)
Severe hypoglycaemia	31 (0%)	32 (0%)	1.614 (0.963–2.709)
Diabetic neuropathy	3,405 (2.5%)	1,968 (2.8%)	1.094 (1.029–1.163)
Diabetic Retinopathy	5,371 (4%)	3,103 (4.4%)	1.099 (1.045–1.155)
Keto-/lactate acidosis	356 (0.3%)	121 (0.2%)	0.697 (0.559–0.862)
Lower limb amputations	807 (0.6%)	408 (0.6%)	0.794 (0.694–0.908)
COPD	14,496 (10.8%)	8,104 (11.4%)	1.03 (0.997–1.064)
Cancer	25,794 (19.1%)	11,160 (15.7%)	0.863 (0.84–0.885)
Major organ specific bleeding	2,025 (1.5%)	577 (0.8%)	0.531 (0.481–0.586)
Bariatric surgery	169 (0.1%)	184 (0.3%)	1.473 (1.184–1.836)
Kidney Transplantation	765 (0.6%)	219 (0.3%)	0.596 (0.505–0.7)
Dialysis	615 (0.5%)	40 (0.1%)	0.152 (0.108–0.209)
Comedications N (%)			

	Dpp4-i	Dapagliflozin	OR (95% CI)
Metformin	65,142 (48.3%)	27,587 (38.8%)	0.635 (0.623–0.648)
Sulfonylurea	12,654 (9.4%)	5,703 (8%)	0.8 (0.773–0.828)
Sulfonamides	3,653 (2.7%)	1,077 (1.5%)	0.666 (0.619–0.716)
Alfa glucosidase inhibitors	479 (0.4%)	119 (0.2%)	0.546 (0.441–0.67)
Thiazolidinediones	471 (0.3%)	312 (0.4%)	1.167 (1.004–1.355)
Other GLDs	11,494 (8.5%)	5,400 (7.6%)	0.843 (0.813–0.873)
Insulin	23,910 (17.7%)	12,400 (17.4%)	0.974 (0.948–1)
ACE inhibitors	30,521 (22.6%)	15,858 (22.3%)	0.974 (0.951–0.997)
ARBs	27,615 (20.5%)	14,878 (20.9%)	1.092 (1.066–1.119)
Beta-blockers	41,508 (30.8%)	25,229 (35.5%)	1.284 (1.256–1.313)
Low/high ceiling diuretics	41,704 (30.9%)	19,614 (27.6%)	1.081 (1.054–1.11)
Aldosterone antagonists	10,819 (8%)	8,326 (11.7%)	1.621 (1.564–1.679)
Thiazide diuretics	4,221 (3.1%)	1,983 (2.8%)	0.964 (0.911–1.02)
Calcium channel blockers	32,750 (24.3%)	15,440 (21.7%)	0.989 (0.966–1.013)
Digitoxin/digoxin	5,012 (3.7%)	2,036 (2.9%)	0.832 (0.786–0.881)
Antiarrhythmics	3,297 (2.4%)	2,236 (3.1%)	1.234 (1.163–1.308)
Statins	80,250 (59.5%)	44,918 (63.1%)	1.106 (1.084–1.129)
Warfarin	616 (0.5%)	294 (0.4%)	1.024 (0.882–1.186)
Aspirin	38,501 (28.6%)	20,588 (28.9%)	1.066 (0.947–1.201)
Purinergic receptor P2Y_12 antagonists	7,158 (5.3%)	3,621 (5.1%)	0.914 (0.874–0.956)
Other antiplatelets	39,406 (29.2%)	20,987 (29.5%)	0.931 (0.827–1.047)
Corticosteroids	20,890 (15.5%)	9,815 (13.8%)	0.992 (0.965–1.021)
Weight loss drugs	5 (0%)	< 5 (%)	1.078 (0.206–4.851)
Lifestyle N (%)			

	Dpp4-i	Dapagliflozin	OR (95% CI)
Tobacco use			
None	105,932 (78.6%)	51,879 (72.9%)	Reference
Past	14,085 (10.4%)	8,550 (12%)	1.094 (1.058–1.131)
Current	14,817 (11%)	10,730 (15.1%)	1.049 (1.019–1.081)
Established ACVD			
Ischemic heart disease	9,145 (6.8%)	4,969 (7%)	1.158 (1.106–1.212)
Ischemic cerebrovascular disease	6,006 (4.5%)	1,808 (2.5%)	0.777 (0.732–0.824)
ACVD Risk factors			
Peripheral arterial disease	2,882 (2.1%)	1,170 (1.6%)	0.777 (0.716–0.842)
Dyslipidaemia	28,445 (21.1%)	13,363 (18.8%)	0.996 (0.968–1.025)
Hypertension	39,783 (29.5%)	16,454 (23.1%)	0.886 (0.861–0.911)

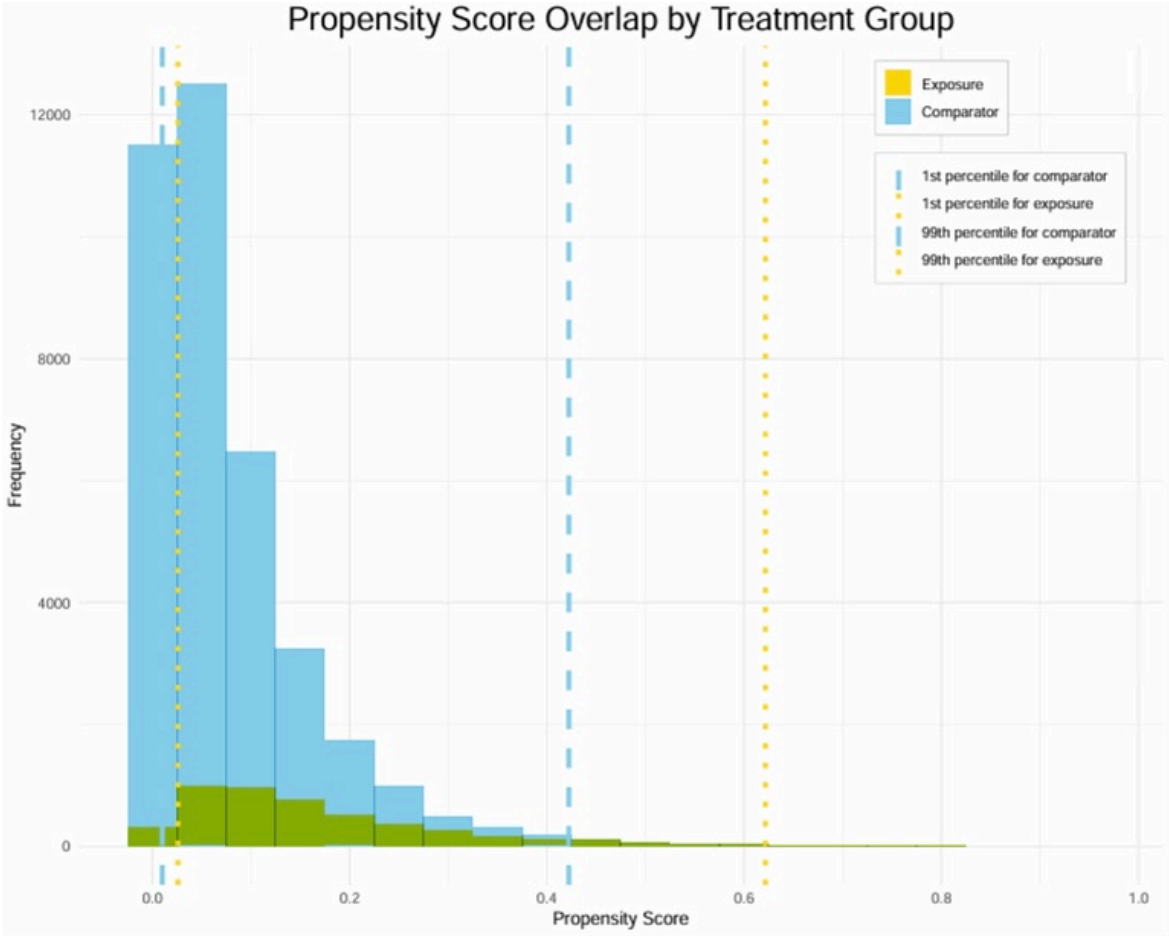
Supplemental Table 7. IPT weights distribution, by treatment arm, before and after truncation at 1 at 99% for CPRD

Characteristic	Dapagliflozin	DPP4-I
Propensity score		
Mean (SD)	0.402 (0.136)	0.316 (0.133)
Q1 – Q3	0.301 – 0.497	0.215 – 0.401
Median	0.394	0.298
Min – Max	0.012 – 0.913	0.008 – 0.889
P1 – P99	0.131 – 0.726	0.080 – 0.666
If truncated, cut-offs	NA	NA
IPTW		
Mean (SD)	0.978 (0.388)	1.002 (0.248)

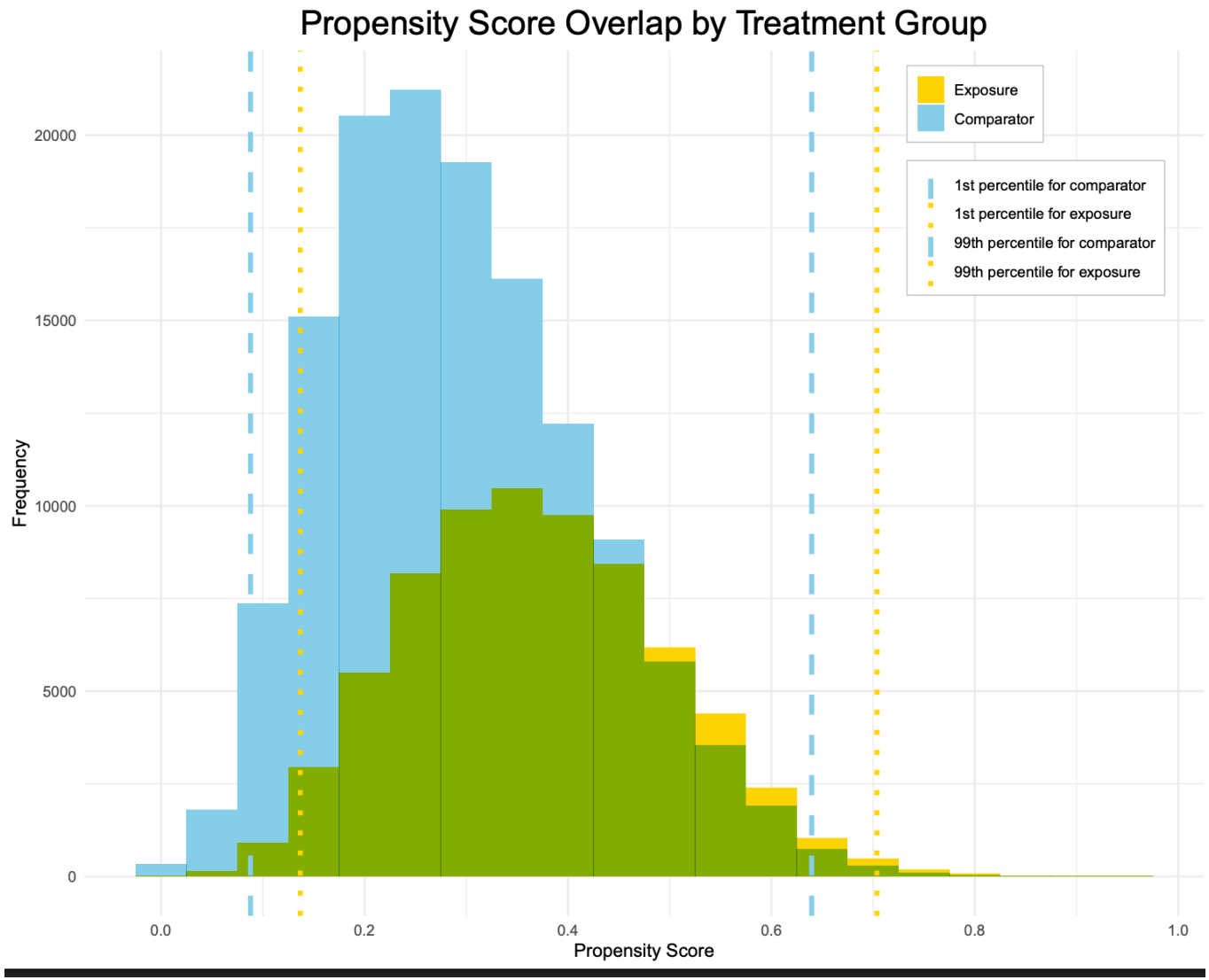
Characteristic	Dapagliflozin	DPP4-I
Q1 – Q3	0.695 – 1.148	0.834 – 1.092
Median	0.876	0.933
Min – Max	0.516 – 2.234	0.660 – 2.234
P1 – P99	0.516 – 2.234	0.711 – 1.962
If truncated, cut-offs	0.516 – 2.234	0.516 – 2.234

Supplemental Table 8. IPT weights distribution, by treatment arm, before and after truncation at 1 at 99% for BIFAP

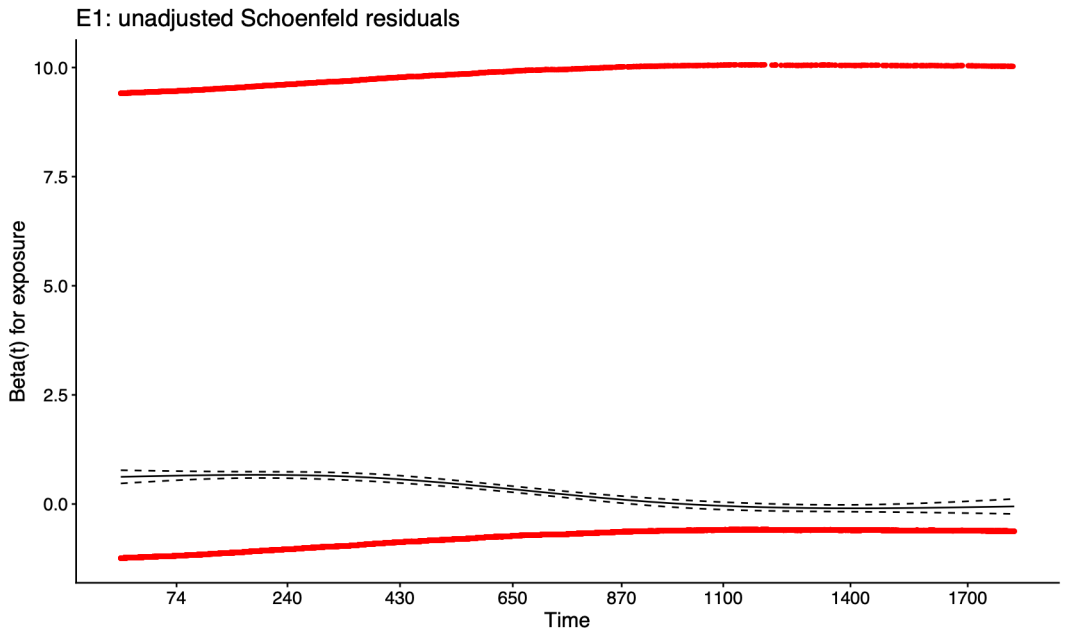
	Dapagliflozin	DPP-4i
Propensity score		
Mean (SD)	0.402 (0.136)	0.316 (0.133)
Q1 – Q3	0.301 – 0.497	0.216 – 0.401
Median	0.394	0.298
Min – Max	0.012 – 0.914	0.009 – 0.883
P1 – P99	0.131 – 0.725	0.080 – 0.666
If truncated, cut-offs	NA	NA
IPTW		
Mean (SD)	0.978 (0.387)	1.002 (0.248)
Q1 – Q3	0.695 – 1.147	0.834 – 1.092
Median	0.876	0.933
Min – Max	0.517 – 2.231	0.660 – 2.231
P1 – P99	0.517 – 2.231	0.712 – 1.961
If truncated, cut-offs	0.517 – 2.231	0.517 – 2.231



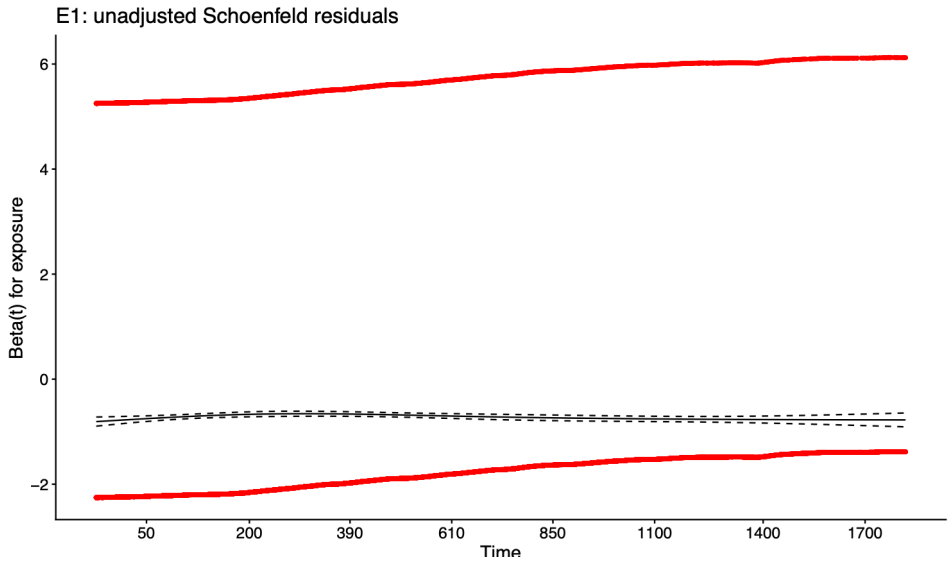
Supplemental Figure 1. Distribution and diagnostics of propensity scores, CPRD



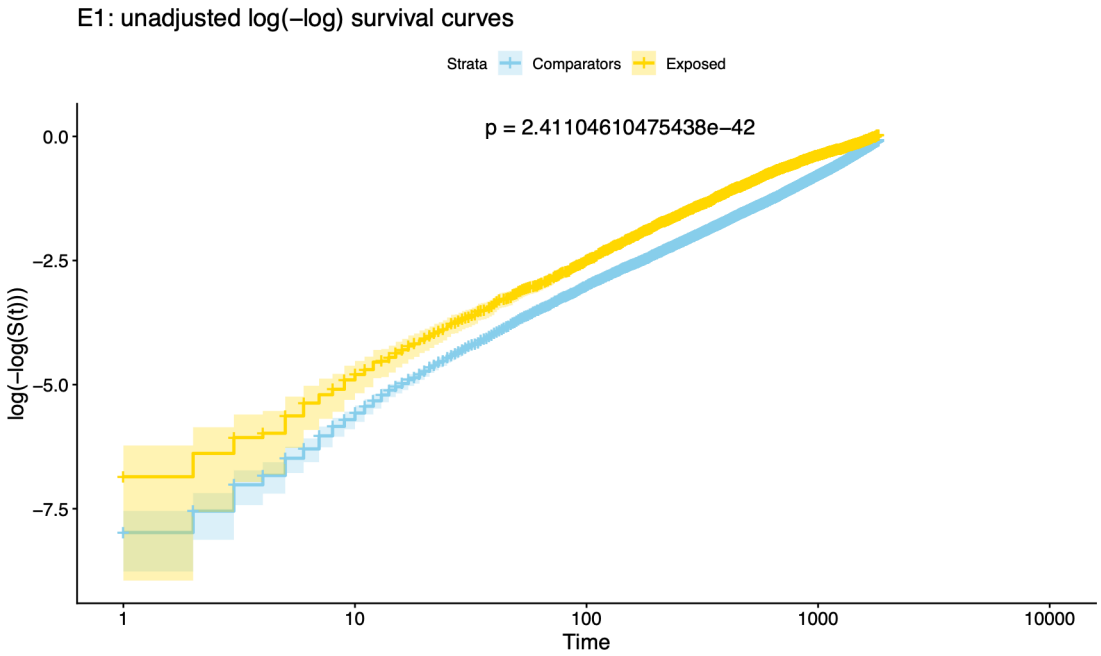
Supplemental Figure 2. Distribution and diagnostics of propensity scores, BIFAP



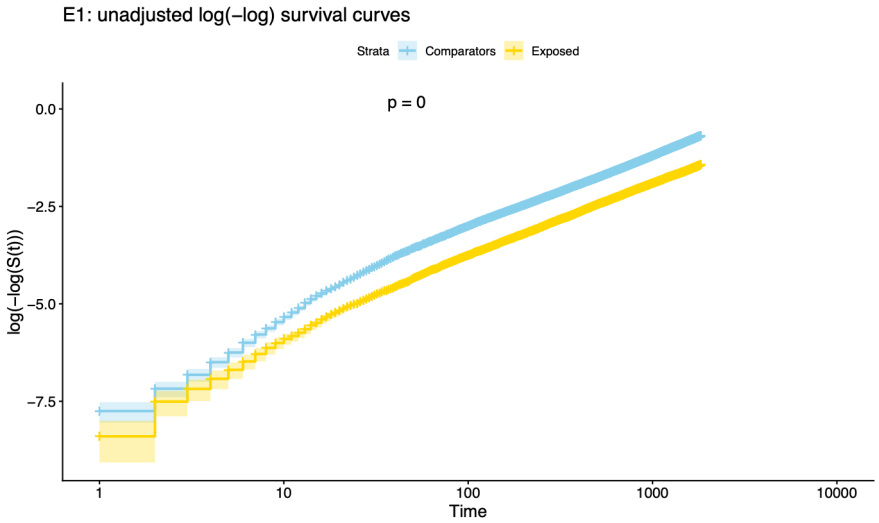
Supplemental Figure 3. Schoenfeld Residuals Plot: Proportional Hazards Assumption for Dapagliflozin vs. DPP-4 Inhibitors (Outcome: MACE), CPRD



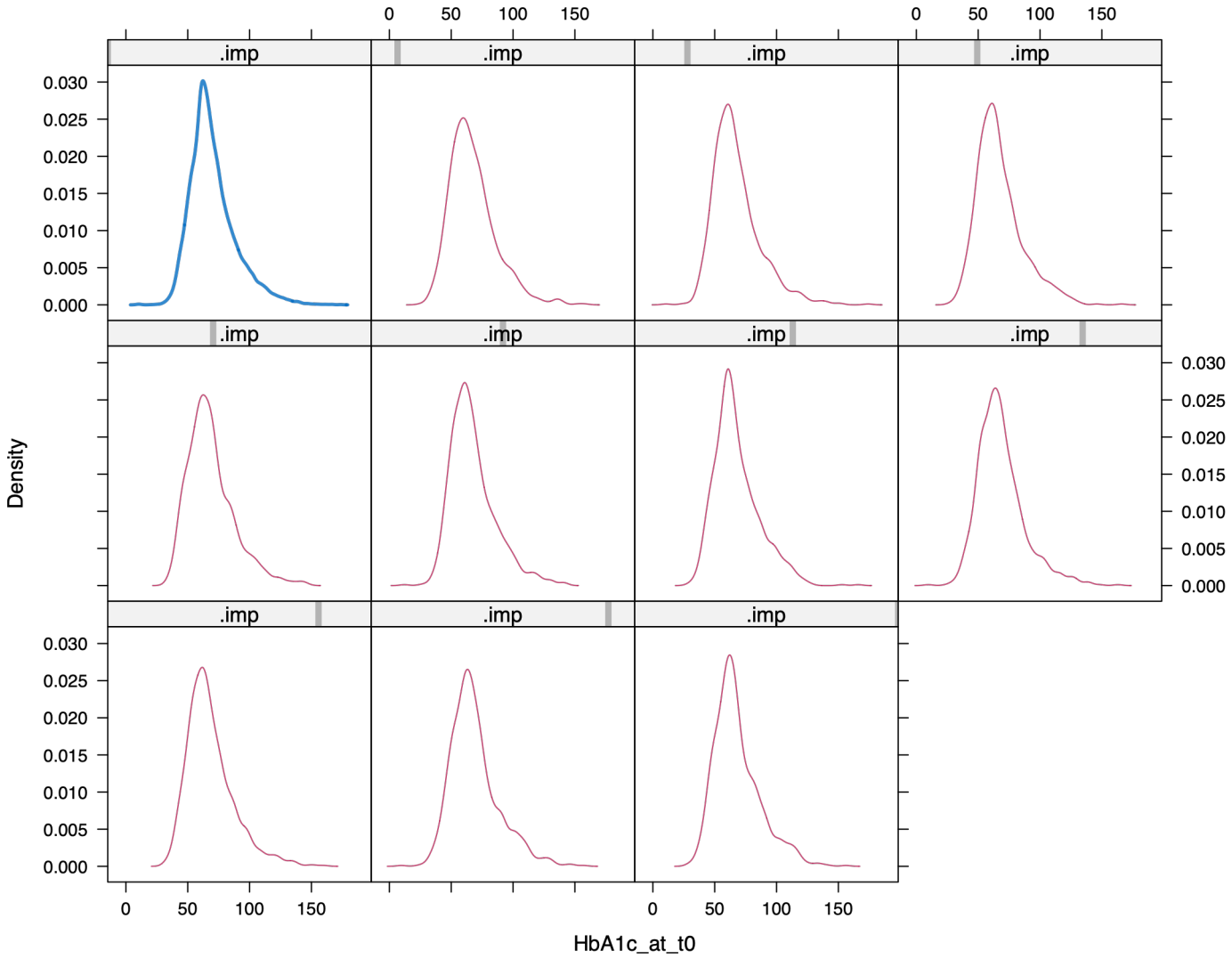
Supplemental Figure 4. Schoenfeld Residuals Plot: Proportional Hazards Assumption for Dapagliflozin vs. DPP-4 Inhibitors (Outcome: MACE), BIFAP



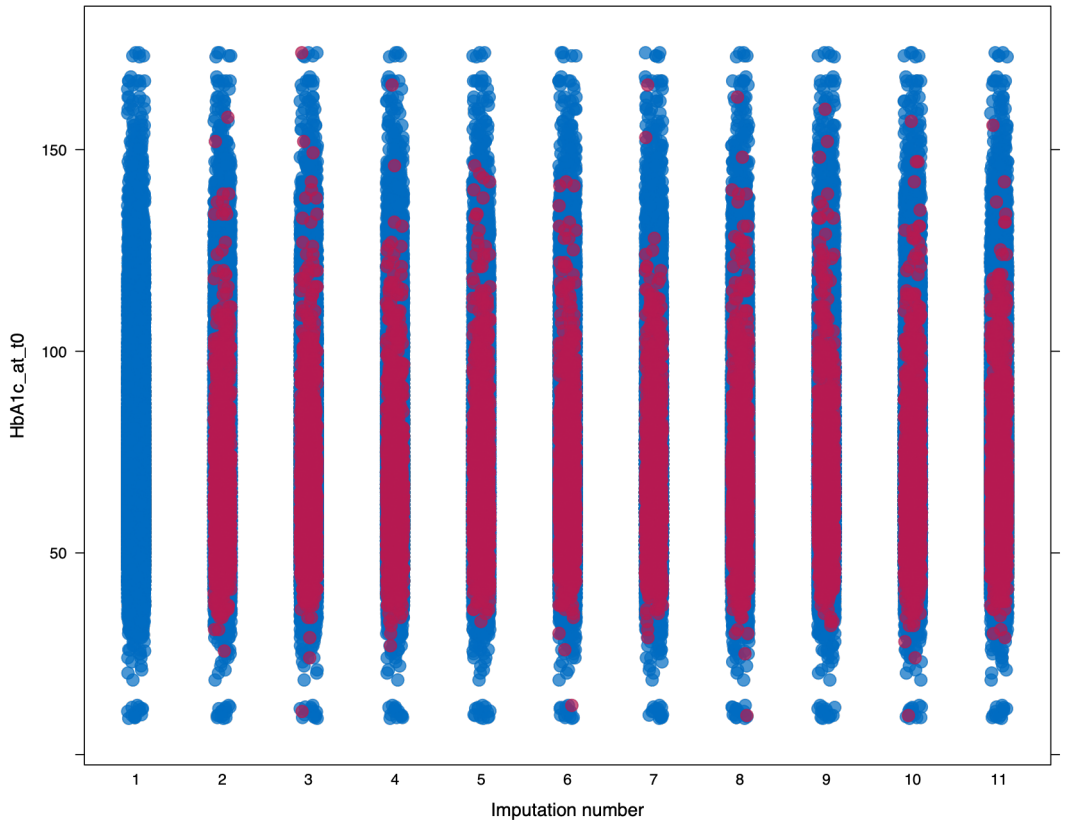
Supplemental Figure 5. Log-Log Survival Plot for Dapagliflozin vs. DPP-4 Inhibitors; Assessment of the Proportional Hazards Assumption for MACE, CPRD



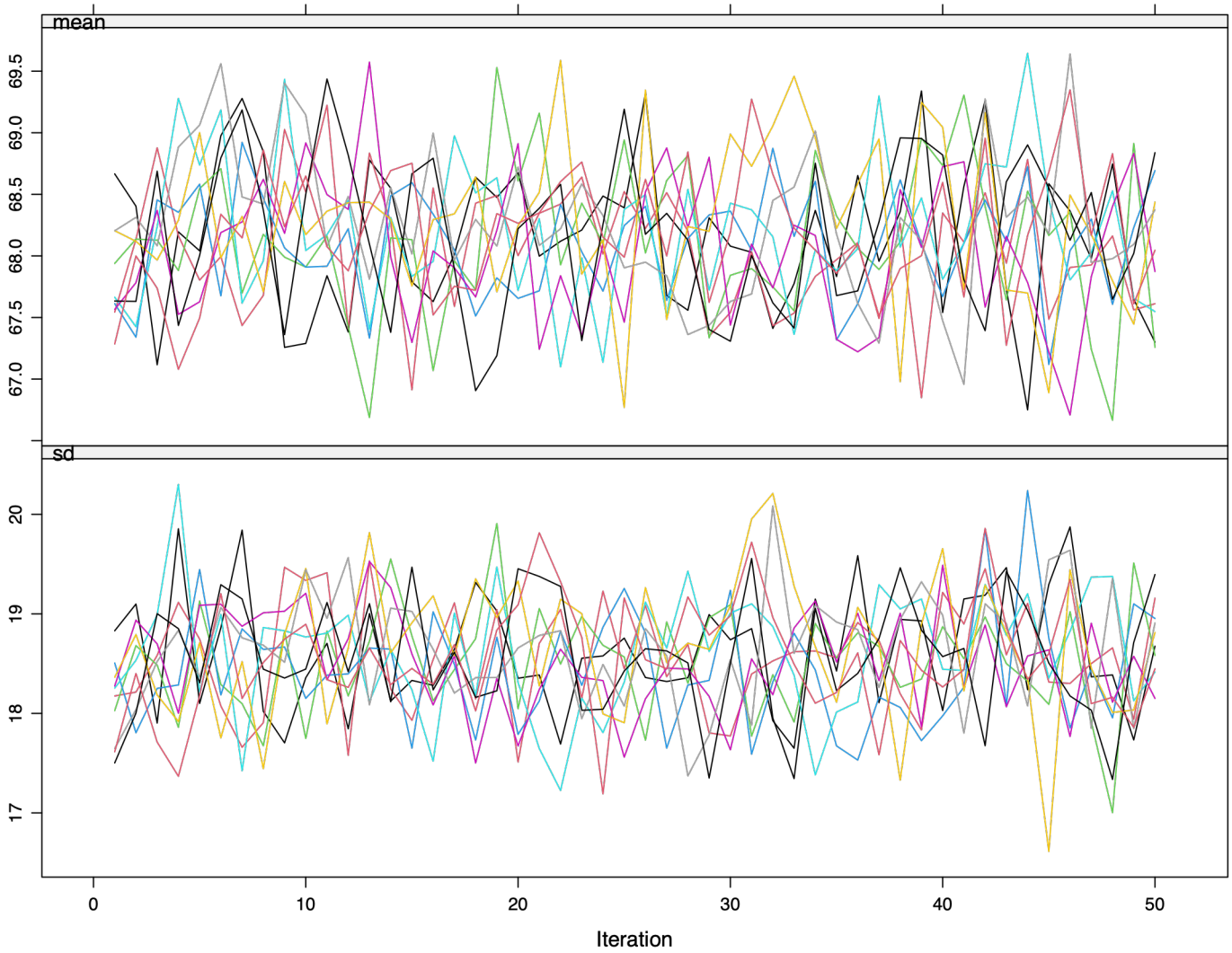
Supplemental Figure 6. Log-Log Survival Plot for Dapagliflozin vs. DPP-4 Inhibitors; Assessment of the Proportional Hazards Assumption for MACE, BIFAP



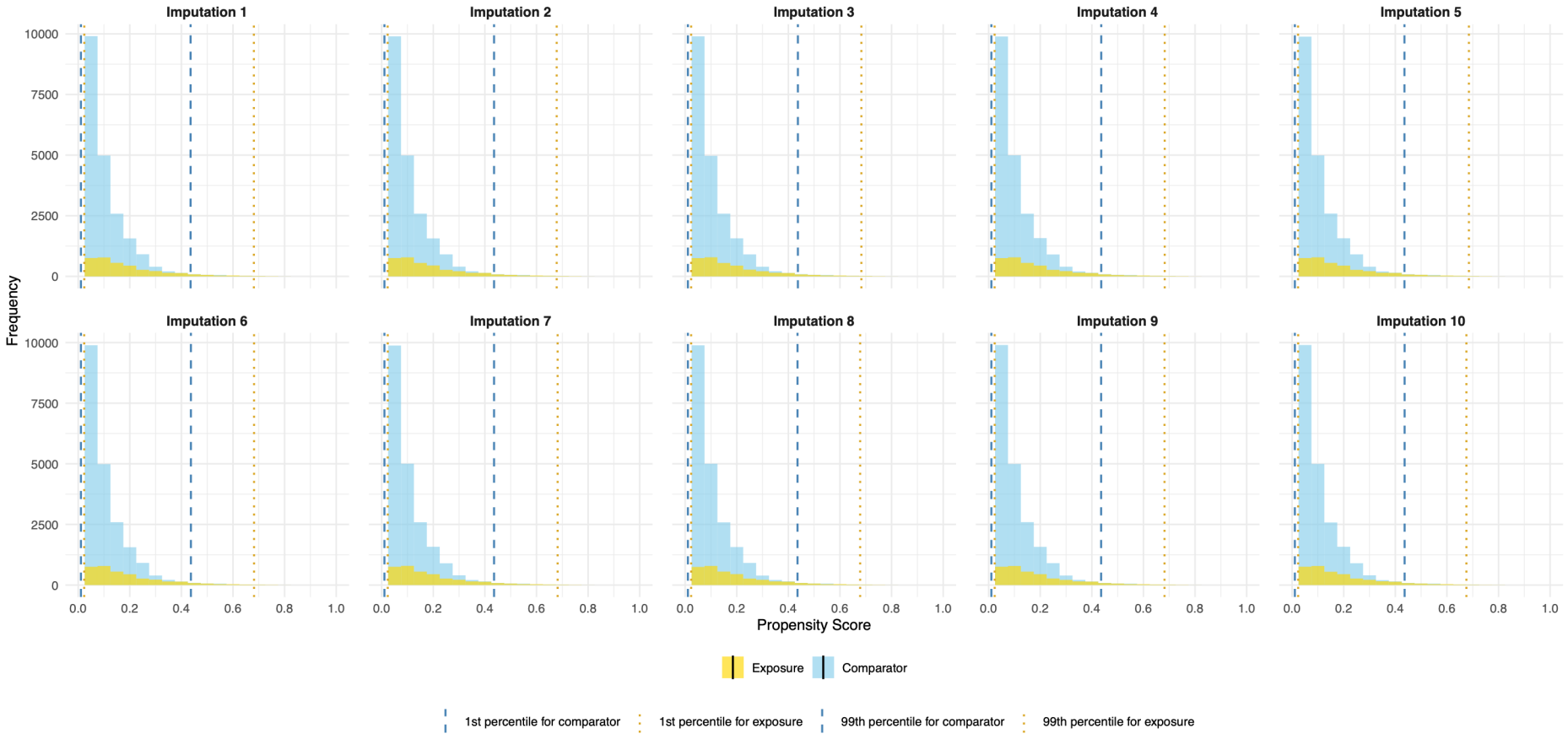
Supplemental Figure 7. Density plot of observed and imputed values of HbA1c at baseline, CPRD



Supplemental Figure 8. Strip plot of observed and imputed values of HbA1c at baseline, CPRD



Supplemental Figure 9. Trace plot of imputation convergence for HbA1c at baseline, CPRD



Supplementary Figure 10. Propensity Score Overlap Across HbA1c Imputations

15.2. *Precision-Based Sample Size Rationale*

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 (1,070), 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%

Supplemental Table 9. % 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

15.3. *Study Meta data (code lists, code books)*

Metadata related to this case study, including the corresponding codelists and codebooks, can be accessed through Zenodo: CS3 Codelist.