Statistical Analysis Plan 18F-MC-GPGN Version 4

The Effect of Tirzepatide versus Dulaglutide on Major Adverse Cardiovascular Events in Patients with Type 2 Diabetes (SURPASS-CVOT)

NCT04255433

Approval Date: 24-Jul-2025

1. Statistical Analysis Plan I8F-MC-GPGN: The Effect of Tirzepatide versus Dulaglutide on Major Adverse Cardiovascular Events in Patients with Type 2 Diabetes (SURPASS-CVOT)

Tirzepatide (LY3298176) for MACE risk reduction

Phase 3, event-driven, multicenter, international, randomized, double-blind, active-comparator, parallel-group study comparing tirzepatide QW up to 15 mg to dulaglutide 1.5 mg QW in patients with Type 2 Diabetes

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I8F-MC-GPGN
Phase 3

Document ID: VV-CLIN-070061

Approved on 24 Jul 2025 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

3. Revision History

Statistical Analysis Plan Version 1, based on the protocol I8F-MC-GPGN(b), was approved on 11 January 2021, prior to the first unblinded review by an Independent Data Monitoring Committee.

Statistical Analysis Plan Version 4 was approved, based on the protocol I8F-MC-GPGN(f) and FDA comments received on 23 July 2025, prior to database lock and unblinding.

Major Revision Summary for I8F-MC-GPGN Statistical Analysis Plan Version 4

Section # and Name	Description of Change	Brief Rationale
4.2, 6.13.3	Replaced a key secondary endpoint rate of change in eGFR per year with change in eGFR from baseline to 36 months.	Addressed FDA feedback on 23 July 2025 regarding SAP V3. 36 months was chosen because all randomized patients had the opportunity to have the visit at 36 months.
4.3, 6.13.5.3	Rate of change in eGFR per year was repurposed as an additional secondary objective.	
6.6	The gamma parameter for the alpha-spending level of the key secondary endpoints was corrected from an incorrect -0.02 to the correct -0.2.	In SAP V2, the gamma was specified as –0.2 as intended, however in SAP V3, it was unintentionally and incorrectly written as –0.02 due to a simple typo when the relevant section was reworded to clarify methodology. There is no change to the methodology itself as specified in SAP V2.

Statistical Analysis Plan Version 3 was approved on 29 April 2025, based on the protocol I8F-MC-GPGN(f).

Major Revision Summary for I8F-MC-GPGN Statistical Analysis Plan Version 3

Section # and Name	Description of Change	Brief Rationale
4.2, 4.3, 6.13.3 and 6.14.2.9	Changed eGFR equation from Creatine-based to Creatine and Cystatin combined-based.	Recommendation 1.2.2.1, Kidney Disease Improving Global Outcomes (KDIGO) 2024 guidelines.
4.2 and 4.3	Changed the definition of specifying patients at high-risk of progression of kidney disease.	KDIGO high-risk or very high-risk cohort was chosen per FDA feedback on 08 March 2024 regarding SAP V2.
6.1	Changed efficacy and safety analyses to include patients regardless of enrollment in sites that were closed due to major GCP violation.	Addressed FDA feedback on 08 March 2024 regarding SAP V2.
6.2	Added the definition of first on-treatment observation period for on-treatment analysis.	Language was added for clarification.

Section # and Name	Description of Change	Brief Rationale
6.2	Added algorithms to determine the uniqueness of recurrent MI and HUA events, and fatal MI or stroke for the analysis of total number of events.	Language was added for clarification.
6.4	Added the method of imputation of missing continuous and binary measures.	To adhere to current Lilly standard for missing data imputation aligned with FDA.
6.6 and 6.13.4	Further clarified the testing scheme considering alpha spending strategy between interim and final analyses, and the graphical testing scheme across primary and key secondary objectives. Descriptions related to above that were previously provided separately and/or duplicated in sections 6.6 and 6.13.4 were consolidated in section 6.6. Contents in section 6.13.4 were removed and merged into section 6.6.	Addressed FDA feedback on 08 March 2024 regarding SAP V2. Note that this is a correction to the description noted by the FDA reviewer. For clarity, this change does not represent change to original plan.
6.13.2.1	In the analysis of imputation of missing follow-up time, a missingness assumption was updated to consider as missing at random for patients due to enrollment at a closure site, i.e., for clearly due to administrative reasons not related to study intervention.	Missing at random assumption deemed to be appropriate for missing data imputation due to site closure for administrative reasons.
6.13.6	Added details on the exploratory objective of indirect comparison of tirzepatide vs placebo.	Refine the analysis to compare tirzepatide to placebo in population similar to SURPASS-CVOT.
6.14.1	Added the list of notable events.	Additional notable events are determined for Patient Narratives
6.14.2.2.2	Added Section for Pancreatic Malignancies.	Analysis is added since pancreatic malignancy is an important potential risk in the tirzepatide core RMP,
6.14.2.6	Added Vitrectomy to the composite endpoint.	To keep consistency between the main study and the substudy regarding the composite endpoint for diabetic retinopathy complications.
6.14.2.7.2	Added category of 8XULN.	Due to the commitment to the FDA regarding patients with ALT/AST levels greater than 8×ULN and those meeting the biochemical criteria for Hy's Law.
6.14.2.8	Added constipation to the GI events.	To align with recent tirzepatide studies.
6.14.2.9	Added definition of ‘persistent’.	Language added for clarification.
Table GPGN 6.4	Added additional criteria ‘≥129 and increase from baseline ≥20’ for systolic blood pressure.	To align with recent tirzepatide studies.
6.14.5	Updated details for diabetic retinopathy substudy specific analyses.	The detailed analysis plans were carefully re-considered in detail.

Section # and Name	Description of Change	Brief Rationale
6.14.5.1	New subsection added for analysis of participants in the Diabetic Retinopathy Substudy.	To describe the unique analytical approach required for this substudy
6.14.5.1	Added definitions for event and censored dates for time to event analyses on substudy specific endpoints.	To clearly define the logic for determining event and censoring dates for substudy specific endpoints.
6.14.5.1	Added the definition of “sustained”	To clarify the criteria of sustained endpoints.
6.14.5.1	Introduced the definition of Diabetic Retinopathy (DR) treatment.	To specify which DR treatments qualify for analysis, as participants receiving treatment are considered to have met the ETDRS2+ or ETDRS3+ endpoint at the relevant time point.
6.14.5.3.1	Updated approach to missing data imputation on substudy specific endpoints.	Replaced retrieved-dropouts imputation approach with multiple imputation under MAR and rule-based event assignment for participants receiving DR treatment, as this provides a more suitable approach for safety analysis.
6.15	Added additional details in PRO analysis	Language added for clarification.
6.16.1	Added subgroup ‘SGLT-2i use at any time on and after baseline’.	This subgroup added to perform for the primary endpoint.
6.16.2	Added subgroup by patients with history of heart failure at baseline.	To pre-specify patients with a history of heart failure as a subgroup of interest relative to the composite endpoint of CV death or HFE.
8	Added additional References.	
Appendix 1	Added table.	Provided table on how to derive the ETDRS scale per person.
Appendix 2	Added statistical analysis for China.	To specify analyses to be performed for participants enrolled in Mainland China and Taiwan, Korea and Japan.
	Throughout the safety section of the SAP made minor changes and reorganization.	Language added for clarification.

Statistical Analysis Plan Version 2 was approved on 16 December 2023 prior to the interim efficacy analysis review by an Independent Data Monitoring Committee held on 30 January 2024.

Major Revision Summary for I8F-MC-GPGN Statistical Analysis Plan Version 2

Section # and Name	Description of Change	Brief Rationale
4.2, 6.13.3 and 6.13.4	Removed hospitalization for unstable angina as a component of the extended composite endpoint: CV death, MI, stroke, hospitalization for unstable angina, and coronary revascularization (MACE-5) endpoint. Accordingly, this outcome is renamed as MACE-4.	Growing evidence suggest that in the era of broad application of high-sensitivity troponin test, the number of individuals diagnosed with unstable angina is decreasing and less reliable due to subjective elements
4.2, 6.13.3 and 6.13.4	Removed MI and stroke from key secondary endpoints.	Results of MI and stroke are expected to be reported as individual key components of the primary composite end points.
4.2, 6.13.3 and 6.13.4	Added rate of change in eGFR per year in patients with chronic kidney disease as a key secondary endpoint.	Emerging evidence supports the eGFR slope as a viable surrogate for kidney failure progression in clinical trials with individuals of chronic kidney disease.
4.2, 6.13.3 and 6.13.4	Changed in the analytical approach of heart failure endpoint to first event of hospitalization or ambulatory visit for heart failure or CV death.	Due to the considerably large number of hospitalizations or ambulatory visits for heart failure as first event, Lilly plans to conduct time to first event analysis without considering repeated heart failure events, as primary analysis
4.3	Added additional composite renal endpoints.	To support assessment for rate of change in eGFR.
5.1, 6.2, 6.13.1 and 6.17	Simplified languages for alpha-levels at the interim and the final with respect to a formal interim efficacy look, by summarizing the details in Section 6.6.	Removed redundancy.
6.1	Exclude data from all patients enrolled in closed sites due to significant GCP violation.	Retaining patient information from sites closed for major GCP violation in the study database could negatively impact data integrity. Lilly plans not considering these data in any pre-specified analyses is in line with similar approach in several CVOTs.
6.6	Clarified alpha spending strategy for primary endpoint and specified it for key secondary efficacy objectives subject to type 1 error rate control, with respective to a formal efficacy interim analysis.	Clarified for primary endpoint, and added for key secondary objectives due to not specified in SAP V1.
6.13.4	Updated graphical testing scheme including layers and alpha level allocation weights changes for key secondary endpoints.	The revised version could increase the likelihood of successful analyzes of individual outcomes.

Section # and Name	Description of Change	Brief Rationale
6.14.2.5	Removal of analyses of hypersensitivity events within or >24 hours from the drug administration	Time of administration relative to hypersensitivity events are not sufficiently collected for analyses.
6.14.2.7.2	Specified details of liver enzyme parameters and those categories.	Clarified comprehensively.
6.14.5	Newly added analysis plan specific for measurements in the retinopathy sub-study.	The sub-study was designed and started after the main study started.

4. Study Objectives

4.1. Primary Objective

The primary objective of the study is to assess efficacy of maximally tolerated tirzepatide dose up to 15 mg once weekly (QW) compared to dulaglutide 1.5 mg QW on time to first occurrence of the composite endpoint of Clinical Endpoint Committee (CEC)-confirmed death from cardiovascular (CV) causes, myocardial infarction (MI), or stroke (MACE-3) when both are added to standard of care in patients with Type 2 diabetes mellitus (T2DM) and high CV risk. All cases in which the causes of death cannot be determined (ie, undetermined) will be included in deaths from CV causes for analyses.

Primary analysis will be an assessment of non-inferiority (NI) of tirzepatide to dulaglutide for MACE-3. Contingent on establishing NI, the superiority of tirzepatide compared to dulaglutide for MACE-3 will be evaluated.

4.2. Key Secondary Objectives Subject to Strong Type 1 Error Rate Control

Contingent on successfully establishing superiority of tirzepatide compared to dulaglutide in terms of the primary endpoint, the following secondary objectives are subjected to the strong control of a Type 1 error rate (see Section 6.13.4).

To assess the superiority of QW tirzepatide, up to 15 mg, compared to dulaglutide 1.5 mg QW for

- time to death due to any cause
- time to CV death
- time to first occurrence of CV death or heart failure (HF) event requiring hospitalization and/or urgent HF visit (HFE)
- time to first occurrence of expanded, composite CV outcome defined as either CV death, MI, stroke or coronary revascularization (MACE-4)
- change from baseline to 36 months in eGFR (CKD-EPI Creatinine-Cystatin Equation 2021) in patients with high or very-high risk chronic kidney disease (CKD) as defined by Kidney Disease Improving Global Outcomes (KDIGO guideline 2025) definition, identified as having 1) eGFR ≥60 mL/min/1.73 m^2 and UACR >300 mg/g, 2) eGFR 45 to <60 mL/min/1.73 m2 and UACR >30 mg/g, or 3) eGFR <45 mL/min/1.73 m^2, at baseline.

The HFEs and MACE-5 (composite of CV death, MI, stroke, coronary revascularization, or hospitalization for unstable angina) added were not designated as key secondary objectives subject to strong Type 1 error rate control in the initial version of protocol, and repositioned in the amended version of protocol I8F-MC-GPGN(c). The incidence of HF events is of special interest. Prior studies in diabetes patients following bariatric surgery have demonstrated that substantial weight loss may reduce the incidence of HF (Sundström et al. 2017). Among various

CV outcomes, the magnitude of reduction was the largest for HF (Aminian et al. 2019). The mechanisms responsible for this potential benefit are not well defined, but substantial weight loss affects cardiac metabolism, function, workload, and anatomy (Aggarwal et al. 2016). Metabolic surgery is also associated with improvements in glycemic control, blood pressure control, and with a lower incidence of other cardiac events (eg, MI and atrial fibrillation), all of which may contribute to the reduction in the risk of HF. Tirzepatide treatment, similar to metabolic surgery, is associated with weight loss, improvement in metabolic conditions like glucose and triglyceride reduction, small blood pressure reduction, and potentially a reduction in MI incidence. All these changes could reduce HF.

MACE-4 is an extended CV outcome measure. Beyond MACE-3, it contains coronary revascularization. Coronary revascularization indicates advanced coronary disease state with high risk for near-term future complications.

4.3. Additional Secondary Objectives Not Subject to Type 1 Error Rate Control

The following objectives are secondary and are not subjected to the strong control of a Type 1 error rate.

To assess the superiority of QW tirzepatide, up to 15 mg, compared to dulaglutide 1.5 mg QW for

- proportion of patients with ≥5, 10, 15 and 20% weight loss from baseline after 36 months
- change after 36 months in
    - weight (absolute and percent changes), body mass index (BMI), and waist circumference
    - hemoglobin A1c (HbA1c) and fasting serum glucose (FSG)
    - blood pressure
    - urinary-albumin-to-creatinine ratio (UACR) (percent change)
    - eGFR (CKD-EPI Creatinine Equation 2021, and CKD-EPI Cystatin C equation 2012), and
    - blood lipids - total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, non-HDL, and triglycerides
- time to first occurrence of
    - MI
    - Stroke
    - coronary revascularizations
    - type of coronary revascularizations: non-elective and other (ie, elective or staged)
    - hospitalization due to UA,

  - composite endpoint of new or worsening nephropathy (composite renal endpoint 1, see definition in Section 6.14.2.9), for overall and patients with high or very-high risk CKD (eGFR is based on CKD-EPI Creatinine-Cystatin Equation 2021),
  - composite renal endpoint 2: a) onset of persistent ≥40% reduction in eGFR, b) ESRD, or c) death due to renal disease, for overall and patients with high or very-high risk CKD (eGFR is based on CKD-EPI Creatinine-Cystatin Equation 2021),
  - composite renal endpoint 3: i) onset of renal failure, defined as initiation of chronic renal replacement therapy (dialysis or kidney transplantation) or persistent eGFR<15 mL/minute/1.73 m^2, ii) death due to renal disease, iii) CV death, or iv) onset of persistent ≥50% reduction in eGFR for overall and patients with high or very-high risk CKD (eGFR is based on CKD-EPI Creatinine-Cystatin Equation 2021), and
  - CV death, MI, stroke, coronary revascularization or hospitalization for unstable angina (UA) (MACE-5).
- cumulative number of CV death, MI, stroke, coronary revascularization or hospitalization for unstable angina (UA) (MACE-5)
- cumulative number of primary composite events of CV death and total (first and recurrent) nonfatal MI(s) and/or nonfatal stroke(s)
- cumulative number of CV deaths and total (first and recurrent) HFEs requiring hospitalization and/or urgent HF visits, and
- rate of change in eGFR (CKD-EPI Creatinine-Cystatin Equation 2021, CKD-EPI Creatinine Equation 2021 and CKD-EPI Cystatin C equation 2012) per year in patients with high or very-high risk chronic kidney disease (CKD) as defined by Kidney Disease Improving Global Outcomes (KDIGO guideline 2025) definition, identified as having 1) eGFR ≥60 mL/min/1.73 m^2 and UACR >300 mg/g, 2) eGFR 45 to <60 mL/min/1.73 m2 and UACR >30 mg/g, or 3) eGFR <45 mL/min/1.73 m^2, at baseline.

4.4. Safety Objectives

Safety objectives include ruling out >30% increase CV risk with tirzepatide in an indirect comparison to placebo by a direct comparison between tirzepatide and dulaglutide in terms of time to first occurrence of any component event of MACE-3.

Safety objectives also include assessing the effects of the maximally tolerated QW tirzepatide dose up to 15 mg compared to dulaglutide 1.5 mg QW for

- incidence of treatment-emergent adverse events (TEAEs), permanent discontinuation of study drug due to adverse events (AEs), and serious adverse events (SAEs)
- incidence of
  - pancreatitis
  - severe or serious gastrointestinal (GI) events
  - any malignancy (including medullary and papillary thyroid cancers)

  - severe hypoglycemic events
  - immune-mediated reactions, including serious allergic and hypersensitivity reactions
  - hepatobiliary events (eg, acute cholecystitis, acute cholelithiasis and drug-induced liver injury)
  - acute renal failure or exacerbation of chronic renal failure
  - diabetic retinopathy complications, and
  - arrhythmias and cardiac conduction disorders
- change from baseline of
  - blood pressure
  - pulse rate
  - lipase
  - pancreatic amylase, and
  - calcitonin.

4.5. Exploratory Objectives

Exploratory objectives include assessing the effects of QW tirzepatide up to 15 mg compared to dulaglutide 1.5 mg QW on

- change of antihyperglycemic drugs
- patient-reported outcomes
  - Ability to Perform Physical Activities of Daily Living (APPADL)
  - Impact of Weight on Self-Perceptions Questionnaire (IW-SP), and
  - 5-Level European Quality of Life Questionnaire (EQ-5D-5L), and
- time to initiation of insulin of >3 months duration for those patients not treated with insulin at the study start.

And, the effects of QW tirzepatide up to 15 mg indirectly compared to placebo on the primary and key secondary outcomes.

5. Study Design

5.1. Summary of Study Design

Study GPGN (SURPASS-CVOT) is a Phase 3, event-driven, multicenter, international, randomized, double-blind, active-comparator, parallel-group study. This study will assess the effect of QW tirzepatide, up to 15 mg, on MACE-3 compared to dulaglutide 1.5 mg QW when added to the standard of care in patients with T2DM with established CV disease and elevated risk for MACE.

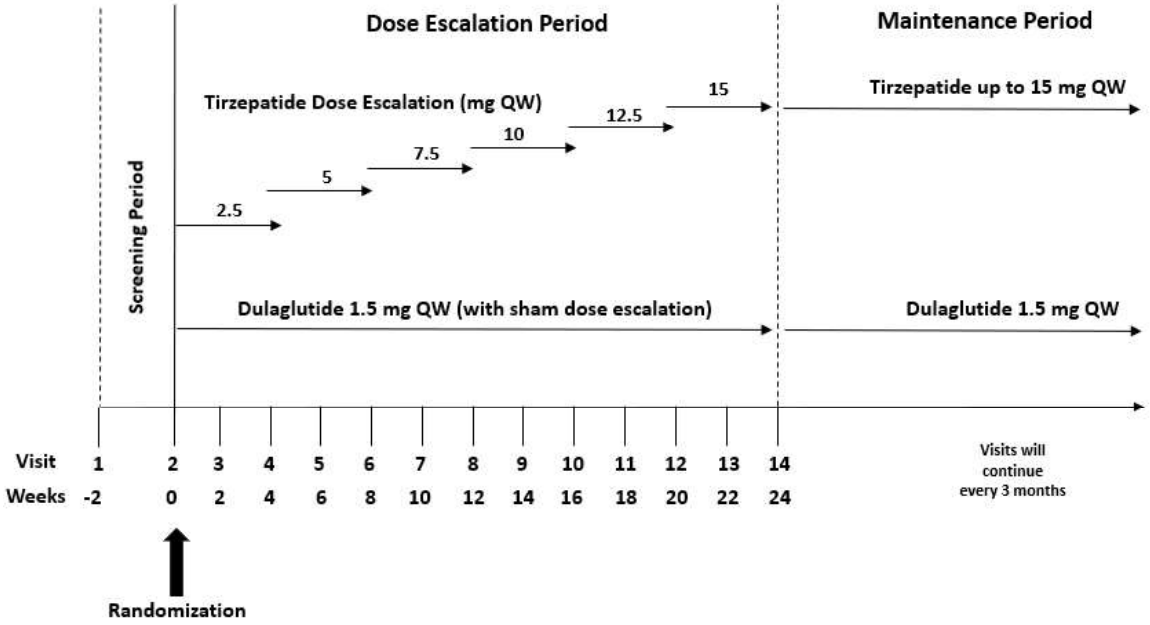
In addition to study drug, all patients should receive standard of care for their diabetes and CV disease.

The study comparison is between tirzepatide and dulaglutide. Assignment to tirzepatide or dulaglutide will be randomly allocated in a 1:1 ratio. Randomization will be stratified by country of enrollment and sodium-glucose cotransporter-2 inhibitor (SGLT-2i) use (Yes/No) at baseline.

Dose Escalation Scheme

The starting dose of tirzepatide is 2.5 mg QW, which is to be escalated at 4 weekly intervals to a maximum of 15 mg QW or to the highest maintenance dose tolerated by the patient. The dose of dulaglutide will be initiated and maintained at 1.5 mg QW. To maintain blinding, a sham escalation of dulaglutide will be employed.

Figure GPGN.5.1 illustrates the study design.



Abbreviation: QW= once weekly.

Figure GPGN.5.1. Illustration of study design for Study I8F-MC-GPGN (SURPASS-CVOT).

At Visit 2, the patient will inject the first dose of study drug at the study site.

For patients randomized to tirzepatide, the starting dose will be 2.5 mg QW with dose escalation every 4 weeks at 2.5-mg increments (2.5 mg to 5 mg to 7.5 mg to 10 mg to 12.5 mg to 15 mg) until the 15-mg dose is reached, or to the highest maintenance dose tolerated by the patient (5 mg or 10 mg).

For patients randomized to dulaglutide, the 1.5-mg QW dose is started at Visit 2 (randomization) and is maintained throughout the study; however, in order to maintain blinding, patients in the dulaglutide treatment group will also be dispensed study drug every 4 weeks during the dose escalation period, following a sham dose escalation schedule to match the tirzepatide treatment group.

To optimize the number of patients who achieve the maximum escalation, both tirzepatide and dulaglutide patients who do not reach the maximum escalation during the initial dose escalation period should have an opportunity to undergo a second dose escalation at Visit 16.

Sample Size, Study Duration, and Visits

Approximately 12,500 patients will be enrolled at approximately 650 sites globally and randomized to 1 of 2 treatment groups: tirzepatide or dulaglutide 1.5 mg. Patients will be randomized in a 1:1 ratio to tirzepatide or dulaglutide using country of enrollment and SGLT-2i use (Yes/No) at baseline as stratification factors. Patients will be followed until at least 1615 patients experience a primary endpoint event, centrally adjudicated as such. This is projected to occur after an average of approximately 48 months of follow-up. The trial may be stopped earlier on the basis of an Independent Data Monitoring Committee (IDMC) safety review or for efficacy at the interim analysis.

Patients will be followed for CV outcomes and other measures every 4 weeks for the first 24 weeks and then followed approximately every 3 months thereafter. A study duration of approximately 54 months is planned, but duration depends on accrual of the requisite number of patients experiencing at least 1 component of the primary endpoint. Consequently, additional visits may be required beyond the planned duration. These additional visits, similar to visits during the maintenance period, will occur every 3 months.

Final Visit

The progress toward the goal of 1615 patients with at least 1 component of the primary endpoint will be monitored during the study based on blinded data. Approximately 3 months prior to the anticipated date of attaining 1615 events, a study close-out period will be declared. During the study close-out period, a final visit will be planned for each patient, with the exception of patients who have died or prematurely discontinued from the study.

If an on-site clinic or telephone final visit with the patient is not possible, investigator site personnel should make every effort to ascertain vital status and capture endpoints by contacting the patient's family or designee/personal contact (eg, friend), family doctor, or attempt to determine vital status via other means (if not prohibited by local laws; eg, national registries/databases, medical records, voter records, and third-party patient locator services). At a minimum, vital status (alive or dead) should be ascertained for all randomized study participants,

including participants who are considered high risk for being lost to follow-up or participants who have withdrawn consent.

Interim Analysis

A single, planned interim efficacy analysis will be performed once approximately 1075 CEC-confirmed MACE-3 endpoints have been accrued to evaluate the need for further continuation of the study.

5.2. Determination of Sample Size

This is an event-driven study. The study will be continued until at least 1615 patients experience 1 or more components of the CEC-confirmed MACE-3 endpoint. This sample size, ie, 1615 patients with MACE-3, provides 90% power to demonstrate the superiority of tirzepatide compared to dulaglutide, relative to MACE-3 risk, at a 2-sided alpha of a 0.0472 significance level if tirzepatide is associated with a 15% reduction in hazard compared to dulaglutide for the MACE-3 endpoint (nQuery + nTerim 4.0 software). The significance level is adjusted to account for an efficacy assessment at an interim analysis. It is assumed that the log-rank test will be used for comparing the distribution of time to first occurrence of MACE-3 between tirzepatide and dulaglutide.

Although the sample size is guided by superiority, the primary study objective is to demonstrate NI of tirzepatide to dulaglutide. An NI boundary of 1.05 will be used, which represents the inverse of the 97.5th percentile of the posterior distribution of the hazard ratio (HR) for dulaglutide versus placebo for the MACE-3 endpoint. The posterior distribution of the HR was derived using a Bayesian, hierarchical meta-analysis of patients with a history of CV disease from 6 completed glucagon-like peptide-1 receptor agonist (GLP-1 RA) cardiovascular outcomes trials (CVOTs), including the REWIND CVOT (Gerstein et al. 2019). Additionally, the NI margin of 1.05 ensures preservation of at least 50% of the dulaglutide efficacy estimated using the Bayesian, hierarchical meta-analysis. The details regarding the justification of the NI boundary of 1.05 is provided in Section 6.7. The accrual of 1615 patients with at least 1 component of MACE-3 provides 90% power to establish NI of tirzepatide to dulaglutide (the upper bound of the 95.3% confidence interval [CI] of the HR, tirzepatide versus dulaglutide, is ≤1.05), provided that tirzepatide is associated with a 10.5% reduction in hazard compared to dulaglutide, relative to the MACE-3 endpoint.

Extrapolated on evidence from previous trials (Zinman et al. 2015; Marso et al. 2016; Hernandez et al. 2018; Gerstein et al. 2019), approximately 35 patients are expected to experience the MACE-3 outcome for 1000 patient-years of exposure in the dulaglutide treatment group. Therefore, approximately 12,500 patients enrolled and followed for an average of 4 years will lead to the accrual of at least 1615 primary endpoints. It is assumed that 2% of the potential exposures will not be realized due to discontinuations from the study and deaths unrelated to CV causes. The number of patients enrolled may be changed based on the observed, blinded, aggregate event rate during the enrollment period. Duration of follow-up may change based on the observed, blinded, aggregate event rates; number of patients enrolled; and lost exposure due to study discontinuations.

6. A Priori Statistical Methods

6.1. Population for Analyses

For purposes of analysis, Table GPGN.6.1 defines the following populations and analysis sets.

Table GPGN.6.1. Analysis Populations/Data Sets

Analysis Data Set/Population	Description
Screened Population	All participants who sign informed consent.
Randomized Population	All patients who are randomly assigned to a study treatment group.
Modified Intention-to-treat (mITT) Population	All randomized patients excluding those who were inadvertently randomized in the study[a].
Modified Intention-to-treat (mITT) Dataset	Data from all randomized patients, excluding those who were inadvertently randomized in the study, during the follow-up period (date of randomization to the date of patient's end of follow-up, see Section 6.2)[a].
Safety Population	All randomized patients who take at least 1 dose of double-blind study drug.
Safety Dataset	Data from all randomized patients who take at least 1 dose of double-blind study drug during the follow-up period (date of randomization to the date of patient's end of follow-up, see Section 6.2).

[a] Several regulatory agencies outside the United States requested that any patient who did not meet all of the eligibility criteria and was inadvertently enrolled must be discontinued from the study. An agency also indicated that they should be excluded from efficacy analysis. The FDA agreed with this approach. Study randomization was completed on Aug 5, 2022 and at the time this SAP version 2 was approved, 129 patients were identified as inadvertently enrolled due to not meeting at least one eligibility criteria and were discontinued from the study.

6.2. General Considerations

Statistical analysis of this study will be the responsibility of Eli Lilly and Company or its designee. All statistical analyses will be conducted with SAS Version 9.4 or higher unless otherwise stated. Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other changes to the data analysis methods described in the protocol, and the justification for making the change, will be described in the statistical analysis plan (SAP) or the clinical study report. Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (eg, few events to justify conducting an analysis). Listings of events will be provided in such situations. Additional analyses of the data may be conducted as deemed appropriate without further changes made to the protocol or SAP, even after the primary or final database locks (DBLs).

In statistical summaries and analyses, patients will be analyzed as randomized. The pairs of "target dataset - analyses" will be as follows, unless otherwise specified.

- Randomized – Summary and listing for patients excluded from mITT, summary for disposition, and important protocol deviations (IPDs).
- Modified Intent-to-treat (mITT) dataset - Summaries for disposition, treatment exposure and compliance, patient characteristics, and concomitant therapy, and efficacy summaries and analyses.
- Safety dataset - treatment exposure and compliance, and safety summaries and analyses.

The last available measurement during Visit 1 to Visit 2, prior to study drug administration, will serve as baseline, unless otherwise specified.

For patients with a final visit where the primary endpoint status is ascertained, the patient's end of follow-up will be defined as the earliest of a) date of death or b) date of the patient's final visit. For patients with documented withdrawal of consent, the end of follow-up will be the date they confirm withdrawal of consent. For patients prematurely discontinuing the study but without documented withdrawal of consent, including those who were discontinued from the study due to site closure due to GCP violation or inadvertently enrolled due to not meeting inclusion criteria or meeting exclusion criteria and discontinued from the study, the patient's end of follow-up will be the latest of

a. the date of the patient's post-early termination visit (ETV)

b. the date of the ETV, or

c. the date of last contact when the primary endpoint status is ascertained prior to date of discontinuation. For patients who have neither a final visit where primary endpoint status is ascertained nor are prematurely discontinued from study, the patient's end of follow-up will be the date of last contact when the primary endpoint status is ascertained.

Measurements at the final visit will be mapped to a closest scheduled visit.

Efficacy and safety measures will be summarized by treatment group as randomized. Summary statistics for continuous measures will generally include sample size, mean, SD, median, minimum, and maximum. Summary statistics for categorical measures (including categorized continuous measures) will include sample size, frequency, and percentages. A single, planned interim efficacy analysis will be performed once approximately 1075 CEC-confirmed MACE-3 endpoints in mITT population have been accrued to evaluate the need for further continuation of the study. Details in alpha-level considering a formal interim efficacy analysis is provided in Section 6.6. Consideration will be given to stop the trial if tirzepatide is superior to dulaglutide. If not, the trial will continue until at least 1615 patients experience at least 1 component of the composite MACE-3 endpoint.

All tests of treatment effects, except for the final efficacy analysis subject to Type 1 error rate control, will be conducted at a 2-sided alpha level of 0.05, and the level of confidence of the 2-sided CI will be 95%, unless otherwise stated.

Listings of any reported endpoints and any safety events occurring after a patient's end of follow-up will be provided. For each patient, time-to-event for a specific event of interest will be the number of days between the date of randomization and the onset date of the event plus 1 day if the patient experiences the event, or the number of days between the date of randomization and the date of the patient's end of follow-up plus 1 day if the patient does not experience the event. If a patient experiences multiple events (eg, multiple strokes), the date of the first event will be used, unless otherwise specified. For time-to-event analyses, the censoring date for a patient is the date of patient's end of follow-up.

Person-years of follow -up for a specific event of interest is defined for each patient as the time-to-event divided by 365.25. The incidence rate per 100 person-years of follow-up is defined by dividing the number of patients who developed the event during the study period by the event-specific total person-years of follow-up (ie, time-to-event defined above) multiplied by 100. The absolute risk difference (ARD) for an endpoint is defined as the difference in incidence rate per 100 person-years between the 2 treatment groups (dulaglutide minus tirzepatide).

Analyses will be performed using data regardless of the status of study drug treatment. Additionally, on-treatment analyses will be performed for selected parameters, including rate of change in eGFR per year, change from baseline in eGFR, HbA1c, weight, BMI, waist circumference, lipids, and blood pressure. These analysis will use data during the first on-treatment observation period, defined as up to the date of last study drug plus 7 days, with the date of last study drug defined as the earlier of the patient permanently discontinuing study drug or interrupting study drug for 3 consecutive months or longer. For the purpose of conducting analysis accounting for all outcome events, first or recurring, the following rules will be applied to determine the uniqueness of events. Two or more CEC-confirmed events will be determined to be one event, that is, the event that led to the chain of events according to following rules. Only this event will be included in the analyses, whereas the other "duplicate" events will be disregarded.

1. An MI is followed by another MI within 30 days, these events will be considered as a single event. A stroke is followed by another stroke within 30 days, these events will be considered as a single event.
2. A coronary revascularization in response to either MI or hospitalization due to UA (HUA) per investigator report in the coronary revascularization form will not be considered a distinct event.
3. MI, HUA or CV cause death occurring within 7 days after a coronary revascularization will not be considered a distinct event.
4. HF event is followed by another HF within 7 days, or CV or undetermined cause death within 30 days, these events will be considered as a single event.

If CEC-confirmed MI(s) or stroke(s) occurred in a patient with a CEC-confirmed CV cause or an undetermined death, it will link the CEC-confirmed MI and/or stroke event to the cardiovascular death as follows.

5. If a patient dies due to CV or undetermined cause, the last MI and/or the last stroke occurring within 30 days prior to death will be considered as fatal.

Not all analyses described in this SAP will necessarily be included in the CSRs. Any analysis described in this SAP and not provided in the CSR would be available upon request. Not all displays will necessarily be created as a "static" display. Some may be incorporated into interactive display tools instead of or in addition to a static display.

6.3. Adjustments for Covariates

The study is stratified by country of enrollment and SGLT-2i use (Yes/No) at baseline. The time-to-event data will be analyzed using a Cox proportional hazard model stratified by SGLT-2i use (Yes/No) at baseline where country will not be included in the model because the number of events could be few in certain countries. For the analysis of other measurements, such as laboratory parameters, country/pooled country, and SGLT-2i use (Yes/No) at baseline, will be used as fixed factors and respective baseline value as a covariate.

6.4. Handling of Dropouts or Missing Data

Sensitivity analyses for the primary efficacy analysis, described in Section 6.13.2.1, will be performed by imputing the missing time to the study end from the date of last contact when the primary endpoint status is ascertained for patients who do not have a study end final visit where primary endpoint status is ascertained.

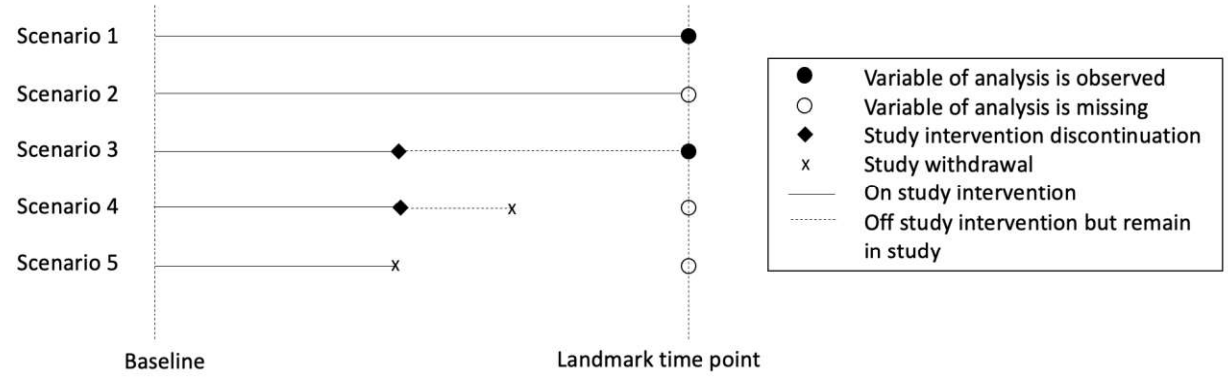
In Section 6.4.1, details of multiple imputation are provided for selected continuous (other than time-to-event) and binary variables, such as change in HbA1c, absolute and percent change in body weight, proportion of patients with <5.7% and ≤6.5% HbA1c and proportion of patients with ≥5, 10, 15 and 20% weight loss. The analysis related to 'hybrid' imputation will also be conducted, based on an analysis of covariance (ANCOVA) model or a logistic regression model. Analysis will be conducted with multiple imputation of missing values at the landmark time point and statistical inference over multiple imputation of missing data guided by Rubin (1987).

The missing data imputation strategy for diabetic retinopathy substudy specific measures is provided separately in Section 6.14.5.

6.4.1. Methods for Multiple Imputation for Continuous and Binary Variables

A landmark time point of 36 months is chosen as most randomized patients are expected to reach 36 months from randomization. Patients were randomized between June 4, 2020 and August 5, 2022, with the close-out period starting on March 24, 2025. Variables related to health outcomes/quality-of-life are collected at baseline, 24 weeks, 24 months as well as the final visit, and lipid parameters are not measured at 36 months, the landmark time point will be 24 months for these variables.

The missing data at landmark time points will be imputed according to Table GPGN.6.2. In general, 5 scenarios are considered regarding the treatment journey of a trial participant relative to the landmark time point, as illustrated below.



For example, data missing due to study withdrawal following permanent discontinuation of study intervention or interruption (Scenario 4) will be imputed using retrieved dropouts (MI-RD), namely using multiple imputation based on data retrieved from participants who permanently discontinued study intervention before the landmark time point or interrupted study intervention at the landmark time point, but with non-missing landmark time point measure from the same treatment group. The overall imputation procedure for creating a single imputed dataset is provided in Table GPGN.6.2. For scenarios 4 and 5A, if there are not enough retrieved dropouts to provide a reliable imputation model (that is, if the model implemented does not converge), an alternative multiple imputation method with reference to the baseline data (that is, a return to baseline approach) will be used.

Table GPGN.6.2 Imputation Procedure

Scenario (treatment/study discontinuation or interruption; missingness at endpoint)	Methods to Handle Missing Values at Endpoint
1. No study intervention discontinuation or interruption; no missing value	N/A
2. No study intervention discontinuation reported before the landmark time point, or intervention interruption at the landmark time point; with missing value	Missing values at landmark time point will be imputed using observed data across all time points (including after the landmark time point) from Scenario 1 and 2 patients under the MAR assumption by treatment group.
3. Study intervention discontinuation reported before the landmark time point, or intervention interruption at the landmark time point; study intervention discontinuation or interruption; no missing value	N/A
4. Study intervention discontinuation followed by study discontinuation, or study intervention interruption at the study discontinuation before the landmark time point; with missing value	Missing values at landmark time point will be imputed under the assumption of multivariate normality for baseline and the landmark time point through MCMC, based upon data only at baseline and landmark time point from Scenario 3 patients (MI-RD) by treatment group.
5. Study discontinuation resulting in study intervention discontinuation before the landmark time point; with missing value	5A: If the study discontinuation other than reason due to enrollment at a closure site and not transferring to another site, the same approach as scenario 4 will be applied to impute missing values by treatment group. 5B: If the study discontinuation is due to enrollment at a closure site and not transferring to another site, i.e., for clearly administrative reasons not related to study intervention, impute the missing data in a manner that reflects what the values would have been without the study discontinuation due to administrative reasons. Specifically, the imputation is conducted in two steps: Step 1: impute the missing data in Scenarios 2, 4, and 5A using the method as described above for each Scenario. Step 2: each imputed dataset (with observed and imputed values from Scenarios 1-5A) will be used to impute the missing values in this Scenario 5B. Only values at baseline and the landmark time point will be included in the imputation model. Since the reference data reflects the distribution of Scenarios 1-4 and 5A in the study population, the imputed values for this Scenario 5B approximates their outcome without study discontinuation due to administrative reasons.

Abbreviations: MAR = missing at random; MCMC = Markov chain Monte Carlo; MI-RD = multiple imputation using retrieved dropouts; N/A = not applicable.

6.5. Multicenter Studies

For the time-to-event analyses, country will not be included in the analysis model because the number of events could be few in certain countries. For other efficacy and safety analyses, such

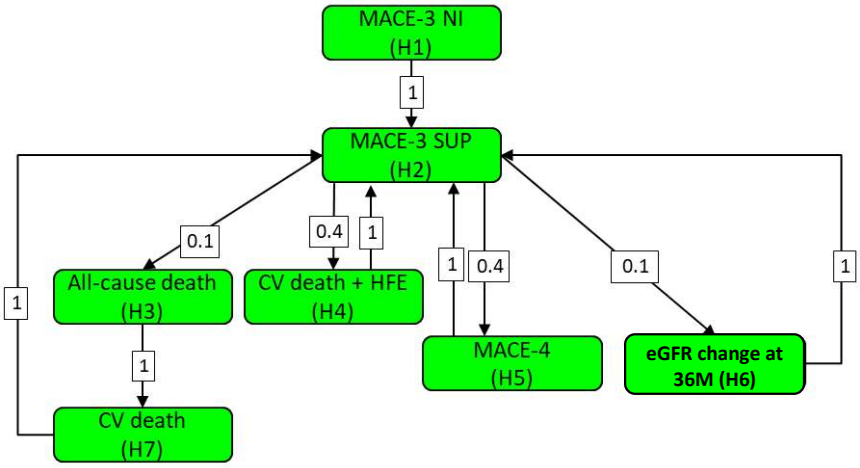
as laboratory parameters, unless otherwise stated, to investigate potential regional influence on parameters, country will be used as a fixed factor in analysis.

6.6. Multiple Comparisons/Multiplicity Adjustments

Multiplicity adjusted analyses will be performed on the primary and key secondary objectives to control the overall family-wise Type I error rate at a 1-sided alpha level of 0.025. The graphical approaches for multiple testing in group sequential trials described in Maurer & Bretz (2013) will be used. This approach strongly controls the Type I error rate across all hypotheses in a group sequential trial setting with a prespecified interim analysis and incorporates alpha spending functions.

The testing strategy is illustrated by Figure GPGN.6.1, where each node represents a hypothesis. The primary endpoint MACE-3 noninferiority is allocated with full alpha initially. Each directed edge is assigned with a transition weight between 0 and 1, representing the fraction of the preserved alpha that is transferred from the hypothesis at the head node to the hypothesis at the tail node, provided the hypothesis at the head node is successfully rejected. Furthermore, the alpha-spending approach will be applied to each hypothesis to control multiplicity at both interim and final analysis.

Specifically, each hypothesis *k* will be tested using an unadjusted one-sided p-value, compared against a rejection boundary defined as nominal significance level $\alpha_k^*(t_k, w_k * 0.025)$, where α_k^* is the function that is derived from alpha spending function $\alpha_k(t, \gamma)$ at information fraction $t = t_k$ and significance level $\gamma = w_k * 0.025$ to obtain the nominal significance level. Alpha spending function α_k and information fraction t_k are specified below; w_k is the weight for obtaining local significance level $w_k * 0.025$ at each step of the sequentially rejective graphical testing procedure.



Abbreviations: NI = non-inferiority; SUP = superiority; CV = cardiovascular; eGFR = estimated glomerular filtration rate; HFE = heart failure (HF) event requiring hospitalization and/or urgent HF visit; MACE -3 = death from cardiovascular or undetermined causes, myocardial infarction, or stroke; MACE-4 = death from cardiovascular or undetermined causes, myocardial infarction, stroke, or coronary revascularization.

Figure GPGN.6.1. Type 1 error control strategy for primary and key secondary efficacy endpoints.

The IDMC will perform a single, planned interim efficacy analysis to evaluate the need for further continuation of the study once approximately 1075 CEC-confirmed MACE-3 events have been accrued. In the interim efficacy review (1075 CEC-confirmed MACE-3 events), it is expected that average duration of follow-up will be approximately 2.5 years.

Consideration will be given to stop the trial if tirzepatide is superior to dulaglutide. If the trial is stopped early for efficacy, the key secondary endpoint analyses will be conducted based on all data up to patient's final visit conducted during close-out period following the early termination announcement. If not, the trial will continue until approximately 1615 patients experience at least one MACE-3 event. Details of alpha spending is provided below:

Alpha-spending level at interim analysis will be determined based on

1. alpha spending function in Hwang et al. (1990) with prespecified parameter, specifically:
   - for the primary endpoint (i.e., MACE-3), gamma = –4.75 is used;
   - for secondary endpoints, gamma = –0.2 is used.
2. For all multiplicity-controlled endpoints, the information fraction at interim will be defined as the proportion of actual number of MACE-3 events accrued by the time of interim analysis, relative to the planned 1615 MACE-3 events at final analysis.

For example, if exactly 1075 MACE-3 events are accrued at interim, the comparison of tirzepatide versus dulaglutide for MACE-3 superiority will be conducted using one-sided significance levels of 0.005 at interim, based on the information fraction 1075/1615.

Alpha-spending level at final analysis will be adjusted based on (1) alpha spent at interim analysis and (2) the actual information fraction updated for each endpoint at final analysis, for example, for MACE-3 superiority, the information fraction is actual number of MACE-3 events in interim analysis divided by actual number of that in final analysis; for CV death, the information fraction is actual number of patients with CV death in interim analysis divided by actual number of that in final analysis.

Unless otherwise specified, there will be no adjustment for multiple comparisons for any other analyses beyond the primary and key secondary objectives.

6.7. Determination of the Non-Inferiority Margin

Per the US Food and Drug Administration (FDA) guidance, "Non-Inferiority Clinical Trials to Establish Effectiveness" the NI margin used in a trial a) can be no larger than the entire assumed effect of the active control in the NI study (M1) and b) represents a clinically acceptable difference between test drug and the active control (M2) (FDA 2016). Justification of NI margin, published in McGuire et.al (2022), is also provided below.

An NI margin of 1.05 for comparing tirzepatide and dulaglutide is chosen because

1. represents the inverse of the 97.5th percentile of the posterior distribution of HR (dulaglutide versus placebo) derived using a hierarchical Bayesian analysis based on the secondary prevention cohort from the REWIND CVOT (Table GPGN.6.3) comparing dulaglutide and placebo and secondary prevention cohorts from several GLP-1 RA CVOTs (Table GPGN.6.4), and
2. an upper bound of a 2-sided 95% CI for the HR (tirzepatide versus dulaglutide) ≤1.05 in a 1615-endpoint CVOT ensures the point estimate of the HR (tirzepatide versus dulaglutide) <0.95; hence, it can be shown using the "synthesis method" outlined in the FDA guidance, "Non-Inferiority Clinical Trials to Establish Effectiveness" that observed HR <0.95 ensures preservation of at least 50% of the dulaglutide efficacy in REWIND CVOT (FDA 2016).

6.7.1. Posterior Distribution of HR (Dulaglutide versus Placebo) in Secondary Prevention

Efficacy of dulaglutide versus placebo was estimated using a Bayesian, hierarchical, meta-analytic model of patients with a history of CV disease, combining REWIND and the other placebo controlled CVOTs of GLP-1 RAs excluding ELIXA. First, exclusion of ELIXA is warranted because the study differed from the other CVOTs in limiting the population to patients who presented within 180 days following an acute coronary syndrome. The mean time from presentation was only 72 days. The pathophysiology and natural history of MACE in patients early after acute coronary syndrome are more related to recurrent thrombosis as opposed to the atherothrombotic events observed in patients with primary or secondary prevention studied in all the other GLP-1 RA CVOTs. In addition, the short-acting pharmacodynamic effect of lixisenatide, the active drug in ELIXA, differs from long-acting GLP-1 RAs evaluated in the other GLP-1 RA CVOTs (Pfeffer et al. 2015). Second, inclusion of the subgroup of patients with

a history of CV disease in the meta-analysis is based on the the fact that the population in GPGN will be limited to these patients.

A Bayesian, hierarchical, meta-analytic model shown below was fitted to the 5 CVOTs in Table GPGN.6.4 and REWIND CVOT in Table GPGN.6.3.

$$y_i \sim N(\theta_i, \sigma_i^2)$$

$$\theta_i \sim N(\theta_c, \tau^2)$$

$$\theta_c \sim N(0, 4^2)$$

$$\tau^2 \sim \Gamma^{-1}(0.001, 0.001)$$

where y_i is the observed log HR of study i and σ_i^2 is the standard error of y_i.

Four independent Markov Chain Monte Carlo chains were used for sampling with a burn-in of 1000 samples on each chain. A total of 400,000 samples were saved and used to back-transform the upper credible limit of the HR (dulaglutide relative to placebo) to the M1 limit. The upper 97.5% quantile of the posterior samples of the HR between dulaglutide and placebo ($UCL_{0.975DulaPbo}$) was used to derive 1.05 of the limit by taking the inverse of the quantity, $1/(UCL_{0.975DulaPbo})$.

This model assumed an overall GLP-1 RA effect relative to placebo and assumed that each individual GLP-1 RA had a log HR randomly distributed around this mean. To arrive at the M1 definition, we inverted the 97.5% percentile of the posterior HR of dulaglutide relative placebo. This value was estimated to be 1.05.

Table GPGN.6.3. REWIND: Number and Percentage of Subjects Reaching the Composite Primary Endpoint of CV Death, Nonfatal MI, or Nonfatal Stroke - Randomization Through Study End, ITT Population

Cohort	Endpoint	Dulaglutide	Placebo	HR
		N, n (%)	N, n (%)	HR (95% CI)
Overall	Primary	4949, 594 (12.00)	4952, 663 (13.39)	0.88 (0.79, 0.99)
Patients with a prior CV event[b,c]	Primary	1560, 280 (17.95)	1554, 315 (20.27)	0.87 (0.74, 1.02)
Patients without a prior CV event[b,c]	Primary	3093, 277 (8.96)	3128, 317 (10.13)	0.87 (0.74, 1.02)

Abbreviations: CI = confidence interval; CV = cardiovascular; ECG = electrocardiogram; HR = hazard ratio (dulaglutide versus placebo); ITT = intent-to-treat; MI = myocardial infarction; N = number randomized; n = number of subjects reaching the primary endpoint; % = 100*(n/N)[a].

[a] Reporting number and percent of subjects with endpoint whether or not the event occurred first.

[b] A prior CV event is defined as a history of MI; a history of myocardial ischemia by a stress test or with cardiac imaging; ischemic stroke; coronary, carotid, or peripheral artery revascularization; unstable angina; hospitalization for unstable angina with ECG changes (new or worsening ST or T wave changes); myocardial ischemia on imaging; or need for percutaneous coronary intervention.

[c] Interaction p-value assessing possible heterogeneity of efficacy with dulaglutide between patients with and without a history of CV disease = .970.

Table GPGN.6.4. Summary of Completed GLP-1 CVOTs for the Meta-Analysis – Primary Endpoints and Associated Hazard Ratio in Patients with a History of Cardiovascular Disease

Study	Drug	Treatment	Number of patients with MACE-3	HR (95% CI)
EXSCEL[a]	Exenatide	Active	722	0.90 (0.82, 1.00)
		Placebo	786	
		Total	1508	
LEADER[b]	Liraglutide	Active	536	0.83 (0.74, 0.93)
		Placebo	629	
		Total	1165	
HARMONY[c]	Albiglutide	Active	338	0.78 (0.68, 0.90)
		Placebo	428	
		Total	766	
SUSTAIN-6[d]	Semaglutide	Active	98	0.72 (0.55, 0.93)
		Placebo	137	
		Total	235	
PIONEER-6[e]	Semaglutide	Active	57	0.83 (0.58, 1.17)
		Placebo	68	
		Total	125	

Abbreviations: CI = confidence interval; CVOT = cardiovascular outcomes trial; GLP-1 = glucagon-like peptide-1; HR = hazard ratio; MACE-3 = death from cardiovascular causes, myocardial infarction, or stroke.

[a] Holman et al. 2017.

[b] Marso et al. 2016b.

[c] Hernandez et al. 2018.

[d] Marso et al 2016a.

[e] Husain et al 2019.

6.7.2. Difference in Cardiovascular Benefit with Tirzepatide and Dulaglutide

It is demonstrated that if the upper bound of the 2-sided 95% CI for the HR of tirzepatide versus dulaglutide in SURPASS-CVOT is ≤1.05, this will result in a clinically acceptable efficacy difference between tirzepatide and dulaglutide by preserving at least 50% of the risk reduction. Therefore, the choice of an NI margin of 1.05 is justified.

If the upper bound of the 95% HR (Tirzepatide versus Dulaglutide) in SURPASS-CVOT with 1615 primary endpoints is ≤1.05, then the observed

$$Ln[HR(Tzp\ vs\ Dula)] + 1.96\ \times \sqrt{4/1615} < Ln(1.05)$$

implying the observed

$$HR(Tzp\ vs\ Dula) < 0.95$$

where, the asymptotic variance of Ln(HR) = 4/D and D = the number of MACE-3 endpoints.

6.7.2.1. Preserved Probability

If an NI boundary of 1.05 in SURPASS-CVOT is met, tirzepatide preserves at least 50% of the efficacy of dulaglutide using a Bayesian method by demonstrating

$$Probability\ (HR(Tzp\ vs\ Pbo) < (HR(Dula\ vs\ Pbo))^{0.5}\ ) > 0.975.$$

$$Probability\ \left(HR(Tzp\ vs\ Pbo) < \left(HR(Dula\ vs\ Pbo)\right)^{0.5}\ \right)$$

$$= Probability\ \left(HR(Tzp\ vs\ Dula) \times HR(Dula\ vs\ Pbo) < \left(HR(Dula\ vs\ Pbo)\right)^{0.5}\ \right)$$

$$= Probability\ \left(HR(Tzp\ vs\ Dula) \times \left(HR(Dula\ vs\ Pbo)\right)^{0.5}\ < 1\right).$$

The posterior distribution of the $HR(Dula\ vs\ Pbo)$ described in Section 6.7.1 will be used as the distribution of *HR (Dula vs Pbo)*.

With a noninformative prior of N(0, ∞) for $Ln[HR(Tzp\ vs\ Dula)]$ and assuming an observed $HR(Tzp\ vs\ Dula) = 0.95$ in SURPASS-CVOT, a posterior distribution of N[Ln(0.95), 4/1600] for $Ln[HR(Tzp\ vs\ Dula)]$ is obtained as follows.

Suppose μ = Ln[HR(tirzepatide versus dulaglutide)] and $\mu \sim N(\mu_0, \sigma^2)$. Let X be the observed Ln[HR(tirzepatide versus dulaglutide)] from the GPGN result.

Then,

$X:\ \mu \sim N\left(\mu, \frac{4}{1600}\right)$. Then $\mu : X \sim N(\mu_1, \sigma_1^2)$, where $\sigma_1^{-2} = \sigma^{-2} + \left(\frac{4}{1600}\right)^{-1}$ and $\mu_1 = \frac{\sigma_1^2}{\sigma^2}\mu_0 + \frac{\sigma_1^2}{\left(\frac{4}{1600}\right)}X$. When $\mu_0 = 0$ and $\sigma^2 \to \infty$, $\sigma_1^2 = \left(\frac{4}{1600}\right)$ and $\mu_1 = X$.

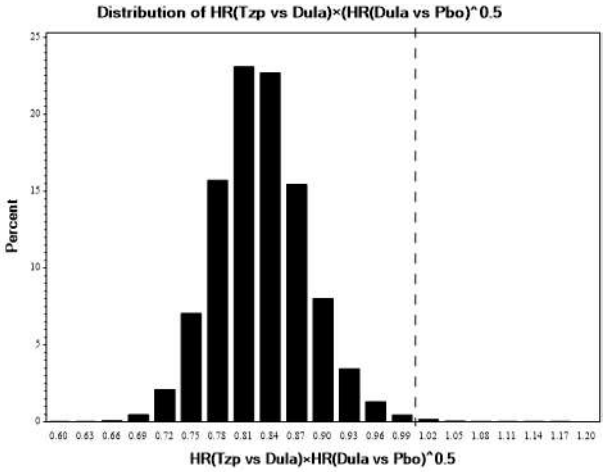
That is,

$$\mu : X \sim N\left(X, \frac{4}{1600}\right) \text{ and hence } \mu : Ln(0.95) \sim N\left(Ln(0.95), \frac{4}{1600}\right).$$

A sample of 99,000 was drawn from each of 2 distributions to simulate HR (Tzp vs Dula)×[HR(Dula vs Pbo)]$^{0.5}$. Figure GPGN.6.2 illustrates the resulting distribution, which showed

$$Probability\left(HR(Tzp\ vs\ Dula) \times \left(HR(Dula\ vs\ Pbo)\right)^{0.5} < 1\right) = 0.997$$

demonstrating an observed value of HR(Tzp vs Dula) ≤0.95 ensures tirzepatide preserving at least 50% of the dulaglutide efficacy.



Abbreviations: Dula = dulaglutide; HR = hazard ratio; Pbo = placebo; Tzp = tirzepatide.

Figure GPGN.6.2. Posterior distribution of HR (Tzp vs Dula) × HR (Dula vs placebo)$^{0.5}$.

6.7.3. Discharging >30% Increase in MACE-3 Risk Compared to Placebo

The 2008 FDA Guidance for Industry, Diabetes Mellitus - Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes directs post-approval CVOTs to be designed to demonstrate that the upper bound of the 2-sided, 95% confidence limit of the HR of MACE-3 for the experimental drug versus placebo to be <1.3. Guidance calls for designing the trial to have 90% power to detect a 30% increase in MACE-3 risk with the new therapy compared to placebo (FDA 2008). Typically, a placebo-controlled trial is continued until 611 MACE endpoints are accrued to discharge this risk by assuming MACE-3 risk is no different in new therapy compared to placebo. SURPASS-CVOT is expected to continue until 1615 MACE-3 endpoints, since the primary objective is to demonstrate tirzepatide is noninferior to dulaglutide. A placebo-controlled trial continued until the accrual of 1615 MACE-3 endpoints provides 90% power to detect a 17.6% increase in MACE-3 risk compared to placebo. Therefore, the upper bound of the 2-sided, 95% confidence limit of the HR (tirzepatide versus placebo) <1.176 is more appropriate to demonstrate that tirzepatide does not increase MACE-3 risk compared to placebo. This choice for the upper bound ensures that the lower bound of the 2-sided, 95% confidence limit of the HR (tirzepatide versus placebo) is <1.00.

Since SURPASS-CVOT is an active-comparator-controlled (dulaglutide) trial, the upper bound must be adjusted to account for the efficacy of dulaglutide (M1). In this case, demonstrating an upper bound of 95% CI for HR, tirzepatide versus dulaglutide, <M1 × 1.176 = 1.050 × 1.176 = 1.235 should satisfy the traditional requirement of demonstrating the upper bound of the 95% CI <1.3 in a 611-primary-endpoint, placebo-controlled trial.

6.8. Patient Disposition

Reasons for screen failure as reported by investigators will be summarized. Listings of randomized treatment assignment (planned treatment) for all randomized patients, patients

discontinued from the study due to inadvertent enrollment and patients enrolled at sites that violated GCP will be provided. Frequency counts of all patients screened will be provided. Frequency counts of all patients randomized as well as GCP violation will be provided by treatment group. Frequency counts and percentages of patients excluded from mITT due to inadvertent randomization and discontinuation from the study, for all randomized patients, will be presented by treatment. Frequency counts and percentages of all patients in mITT, as well as vital status availabilities and those receiving at least 1 dose of study drug, will be presented by treatment. A listing of patients not receiving study drug and a listing of final study disposition will be provided for ITT population. Frequency counts and percentages of patients completing the study, etc. including reason for premature discontinuation and vital status, will be presented by treatment groups for mITT populations. The HR and its 95% CI of time from randomization to premature discontinuation from study will be provided. The estimate of the cumulative curve of premature discontinuation from the study over time will be obtained using the Kaplan-Meier (KM) method. The confirmation method of primary endpoint status ascertainment during the close-out period will be presented by treatment groups.

6.9. Patient Characteristics

Listings of patient demographics will be provided. Demographic, medical history including qualifying CV entry criteria, and concomitant illness will be summarized by treatment group.

History of myocardial infarction, stroke, coronary revascularization, and heart failure will be identified based on from multiple CRFs.

6.10. Concomitant Therapy

Concomitant therapy at baseline will be determined as the medication being taken on the date of randomization. Specifically, for SGLT-2i reported in the baseline characteristics summary and used in analyses, such as a subgroup analysis and a stratified Cox proportional hazard model, patients defined as on SGLT-2i at baseline are patients who initiated on or before the day after randomization and did not stop prior to that day. Similarly, when performing analyses by other concomitant therapies at baseline, such as hypoglycemia analyzed by status of baseline insulin and/or SU use, the same definition as for SGLT-2i will apply.

Antihyperglycemic Therapy

Screening, baseline, post-baseline, and change of antihyperglycemic therapy will be summarized by type and by treatment.

In addition, time to initiation of the following will be analyzed; relevant HRs and 95% CIs will be provided:

- insulin of >3 months duration for those patients not treated with insulin at baseline, and
- SGLT-2i for those patients not using at baseline.

The estimate of the cumulative event curve over time will be obtained using the KM method, where the start date of the drug for patients who use at baseline will be replaced with the randomization date plus 1 day.

Other Concomitant Therapies of Interest

Screening, baseline and post-baseline concomitant therapy, respectively, will be summarized by type and by treatment. Concomitant medications of interest include the following groups of medications:

- antihypertensive therapy, and
- lipid lowering- therapy.

6.11. Treatment Exposure and Compliance

Listing of patients randomized but not receiving study treatment will be provided, if applicable. The listings will include patient identification (ID), randomized treatment group, and the reason for not receiving study treatment.

Frequency counts and percentages of patients by patients' status, e.g., on-treatment (by dose level for tirzepatide group), off-treatment, dead, withdrawal of consent, and lost to follow-up at selected visits will be summarized.

Exposure Summary

Summaries and proportions of patients for the following time categories (eg, yearly) will be provided by treatment group:

- *follow-up time* (defined as time in months between the date of randomization to the date of the end of follow-up plus 1 day), and
- *exposure time on study treatment* (defined as from date of first dose of study treatment to date of last dose of study treatment plus 7 days, removing interruption period) will be provided by treatment group. Summary and proportion of patients by percent category (eg, every 10%) for proportion of exposure time on study drug relative to follow-up time ([exposure time on study treatment/follow-up time] × 100%).

Presentation in months will be derived as study days divided by 30.4375.

Person-Years

The total person-years of study follow-up by treatment group will be calculated as the total number of days between the randomization date and the date of the end of follow up plus 1 day divided by 365.25.

Interruption, Discontinuation, and Compliance

Frequency counts and percentages of patients temporarily interrupting study drug, and prematurely and permanently discontinuing study drug, overall and by reasons, will be summarized by treatment group. The KM method will be used for illustrating the time profile of

the cumulative proportion of patients discontinuing study drug over time by the treatment group. *Study drug compliance* will be defined as

$$\frac{(total\ number\ of\ autoinjectors\ dispensed - total\ number\ of\ unused\ autoinjectors\ returned)}{number\ of\ weeks\ from\ date\ of\ first\ dose\ to\ date\ of\ last\ dose,\ \ excluding\ interruption\ period} \times 100.$$

Summary and proportions of patients of exposure and compliance will be reported by treatment group.

6.12. Important Protocol Deviations

Important protocol deviations will be specified in the trial issues management plan. A listing and a summary of IPDs by treatment will be provided.

6.13. Efficacy Analyses

6.13.1. Primary Efficacy Analysis

The primary outcome will be the occurrence of the CEC-confirmed MACE-3. All cases in which the causes of death cannot be determined (ie, undetermined) will be included in deaths from CV causes for analyses. The primary endpoint will be the time from randomization to the first occurrence of any component of the primary outcome. The primary analysis will be a Cox proportional hazard model stratified by SGLT-2i use (Yes/No) at baseline with treatment as a fixed effect. The HR, with its CI and p-value, will be provided in the primary analysis model.

The final efficacy assessments will be guided by a final alpha level (ie, 100 x [1 - final alpha level]%) CI of a HR (tirzepatide versus dulaglutide) for the primary endpoint. Details in alpha-level considering a formal interim efficacy analysis is provided in Section 6.6. To account for multiple comparisons related to the primary outcome conducted at the final analysis, the following analyses will be conducted hierarchically in the order specified.

Discharging CV Risk Increase Compared to Placebo

- The first assessment is conducted to discharge the FDA regulatory requirement to rule out a >30% increased CV risk with tirzepatide compared to placebo by a direct comparison between tirzepatide and dulaglutide. The requirement calls for demonstrating that the upper bound of the 95% CI of the HR (test drug versus placebo) is <1.3 in a CVOT. Typically, this requirement is discharged using a CVOT with 611 primary endpoints that provide 90% power to discharge >30% increase in hazard (ie, upper bound of the 95% CI of the HR is <1.3) when HR in truth is 1. A CVOT comparing the experimental agent to placebo with 1615 primary endpoints provides 90% power to discharge a >17% increase in hazard (ie, the upper bound of the 95% CI of the HR is <1.17) when HR in truth is 1. Since the comparison is to dulaglutide, if the upper bound of the 95.3% CI of the HR (tirzepatide versus dulaglutide) = the HR (tirzepatide versus placebo) × the HR (placebo versus dulaglutide) <1.17×1.05 = 1.23 is met, then the regulatory requirement is considered met. See Section 6.7 for the justification of the choice of 1.05 as a conservative estimate of HR (placebo versus dulaglutide).

Demonstrating Effectiveness versus Placebo

- Contingent upon successfully demonstrating the steps above, NI of tirzepatide to dulaglutide will be assessed by inspecting the upper bound of the 2-sided, 95.3% CI of the HR (tirzepatide versus dulaglutide), and noninferiority will be established if that upper bound is ≤1.05. When establishing noninferiority to dulaglutide, the demonstrating effectiveness of tirzepatide relative to MACE risk reduction versus placebo would be declared.

Superiority over Dulaglutide

- Contingent on establishing NI of tirzepatide to dulaglutide, superiority of tirzepatide would be claimed if the upper bound of the 2-sided, 95.3% CI is <1.

The 95% CI, in addition to adjusted CI, e.g., 95.3% CI, will also be presented from the Cox model for the primary endpoint. The estimate of the cumulative event curve over time will be obtained using the KM method. Counts and proportions of patients who experience a primary endpoint event will be calculated. The total person-years of follow-up for the primary endpoint, the incidence rate per 100 person-years of follow-up for the primary endpoint, and the ARD for the primary endpoint will be provided.

6.13.2. Sensitivity Analyses of the Primary Outcome

6.13.2.1. Imputation of Missing Follow-Up Time

The purpose of the analysis is, *under the hypothetical situation in which if all patients had been followed until the end of the study, and of those, patients who the primary endpoint was not ascertained had not taken the study drug after censoring date*, to estimate the latent number of patients who would have experienced a primary endpoint along with the HR, with its 95% CI and p-value. In all mITT patients, the simulated number of patients who would have experienced a primary endpoint and the estimated HR, with its 95% CI and p-value, will be presented.

In the mITT dataset, the missing follow-up time for patients for whom the primary endpoint status was not ascertained other than CEC-confirmed non-CV cause death, for reasons other than due to enrollment at a closure site and not transferring to another site, will be imputed by the estimated event-rate model for the primary endpoint from the “retrieved dropouts”. A single, imputed, complete dataset will be generated via following steps.

1. Retrieved dropouts are defined as patients who prematurely and permanently discontinued study drug because of a reason other than death, were not prematurely discontinued from the study, and did not experience a primary endpoint before discontinuing study drug in the same treatment group.

2. The event-rate model will be estimated for the time from study drug discontinuation to the first primary endpoint among retrieved dropouts by treatment group. A piecewise exponential hazards model may be used to model the event rate. The final model, including intervals such as [0, 0.5), [0.5, 1) and [1, infinity) years, will be made prior to unblinding by reviewing interim blinded data distribution. When/if the event-rate functions will not be fitted to the treatment-wise distribution, it may be necessary modify the model after unblinding.
3. Generate missing follow-up time from the censoring date to the end of the study (last day of the study close-out period) for patients for whom the primary endpoint was not ascertained other than CEC-confirmed non-CV cause death according to the estimated model.
4. The simulated time-to-event from the randomization will be given as the sum of the observed time to censor from randomization and the generated missing follow-up time.
5. If the simulated time-to-event is shorter than the duration to the end of study from the randomization, then the patient is considered as to have "experienced" a primary endpoint at the simulated time; if not, the patient is considered as to have been censored at the time of the end of the study.
6. The complete dataset will be obtained by combining the imputed dataset for patients for whom the primary endpoint was not ascertained and the dataset for remaining patients for whom the primary endpoint status was confirmed.

For patients whom the primary endpoint was not ascertained due to enrollment at a closure site and not transferring to another site, i.e., for clearly due to administrative reasons not related to study intervention, the missing follow-up will be imputed as follows:

A. The event-rate model will be estimated in a similar manner as in step 2, except for the time from "randomization" to the first primary endpoint among "all mITT patients" by treatment group.
B. Generate missing follow-up time from the censoring date to the end of the study (last day of the study close-out period).
C. Perform the same as steps 4 to 6.

Iterate (eg, 1000 times) steps 3 to 6 and steps B to C, and obtain the multiple complete datasets for the number of iterations; where, in each iteration of multiple simulations, the different parameters of the estimated model (eg, lambda of scale parameter of the exponential distribution) according to a proper distribution (eg, Bayesian multiple imputation approach) will be used.

The same Cox model as the primary efficacy analysis will be applied for each complete dataset to obtain the parameter estimates (ie, log[HR] and its standard error). The summarized estimates will be obtained using Rubin's formula, and then the scale of back transforming to the HR, its 95% CI and p-value will be presented.

6.13.2.2. Tipping Point Analysis

The purpose of the analysis is, *under the hypothetical situation in which if all patients had been followed until the end of the study, and of those, patients who the primary endpoint was not ascertained had not taken the study drug after censoring date*, to estimate the level of increase of the estimated latent event-rate during the missing follow-up of tirzepatide-treated patients that may alter tirzepatide efficacy.

1. Similar to 6.13.2.1, it's to obtain the summarized estimates 95% CI of HR and p-value, by replacing event rate (hazard rate) in each interval for tirzepatide group in steps 3 and A with increased event rate by a specific ratio, e.g., 5%.
2. Iterate #1 by slightly changing an increased ratio until finding the level of risk increase of tirzepatide-treated patients that changes the conclusion.

6.13.2.3. On-Treatment Analysis

Another sensitivity analysis of the primary analysis will be performed by replacing the censoring date for a patient with an earlier date between the end date of the first on-treatment observation period plus 30 days or end of follow-up for the patient, whichever is earlier.

6.13.3. Secondary Efficacy Analyses Subject to Type 1 Error Rate Control

Analyses will be conducted in terms of the following endpoints related to key secondary objectives subject to Type 1 error rate control:

- time to death due to any cause
- time to CV death
- time to first occurrence of CV death or HF event requiring hospitalization and/or urgent HF visit (ie, HFE), and
- time to first occurrence of the expanded, composite CV outcome defined as MACE-4.
- change from baseline to 36 months in eGFR (CKD-EPI Creatinine-Cystatin Equation 2021) in patients with high or very-high risk CKD

Time to event analysis described in Section 6.13.5.2 will be applied. Change from baseline to 36 months in eGFR will be analyzed based on an analysis of covariance (ANCOVA) model with treatment, country/pooled country, and SGLT-2i use (Yes/No) at baseline as fixed factors and baseline as a covariate. Missing value at 36 months will be imputed as specified in Section 6.4.1.

Of note, while dulaglutide in the US label is only indicated for MACE risk reduction among primary and key secondary endpoints, neither dulaglutide nor other GLP1 RA have shown any adverse effect on risk of all cause death, CV death, hospitalization for HF, coronary revascularization, or GFR progression. The available evidence supports that GLP1-RA's either statistically significantly lower or numerically lower risk of these components compared to placebo (Sattar et.al. 2021; Bortos et.al. 2023; Perkovic et.al. 2024).

6.13.4. Type 1 Error Rate Control Strategy for Primary and Key Secondary Efficacy Analysis

Details of type 1 error rate control strategy for primary and key secondary efficacy analysis in a group sequential trial setting with a prespecified interim analysis and incorporating alpha spending function are provided in Section 6.6.

6.13.5. Other Secondary Efficacy Analyses

6.13.5.1. CEC Adjudication

A summary of the details of CEC adjudication for each event will be provided. The contingency table between investigator-report and adjudicator-assessed will be provided. The time to first occurrence of investigator-reported MACE-3 and each component will be analyzed in the same manner as the primary endpoint analysis model, accordingly, presenting the HR, with its 95% CI and p-value.

6.13.5.2. Event-Related Measures

Time-to-First-Event Analysis

The HR, with its 95% CI and p-value, will be provided in the same manner as the primary endpoint analysis model. The estimate of the cumulative event curve over time will be obtained using the KM method. Counts and proportions of patients who experience an event will be calculated. The total person-years of follow-up for the specific event, the incidence rate per 100 person-years of follow-up for the specific event, and its ARD will be provided. Each component for the composite endpoint will also be analyzed. The time to first occurrence of the following events will be analyzed:

- MI
- Stroke
- coronary, carotid, or peripheral revascularizations, individually and as composite
- type of coronary revascularizations: non-elective and other (ie, elective or staged)
- hospitalization due to unstable angina (UA)
- composite endpoint of new or worsening nephropathy (see definition in Section 6.14.2.9)
- CV death, MI, stroke, coronary revascularization or hospitalization for unstable angina (UA) (MACE-5)
- composite renal endpoints (see definition in Section 4.3)

Peripheral revascularization is defined as the location of revascularization, which is in the femoral, popliteal, or iliac artery.

Cumulative Number of Events

Counts, the rate per 100 person-years of follow-up, and the total person-years of follow-up will be provided. The estimate of the cumulative mean function curve over time will be plotted using

the non-parametric model (Ghosh and Lin 2000) with treatment and SGLT-2i use (Yes/No) at baseline as fixed effects, the rate ratio with its 95% CI and p-value will be provided from the semi-parametric proportional rates model (abbreviated as the LWYY model) (Lin et.al. 2000) with treatment and SGLT-2i use (Yes/No) at baseline as fixed effects for following recurrent events:

- cumulative number of CV death, total (first and recurrent) non-fatal MI(s), total (first and recurrent) non-fatal stroke(s), total (first and recurrent) coronary revascularization(s) or total hospitalization for unstable angina(s) (UA(s)) (MACE-5)
- cumulative number of primary composite events of CV death, total (first and recurrent) non-fatal MI(s), and/or total non-fatal stroke(s)
- cumulative number of CV deaths, total (first and recurrent) HF events requiring hospitalization, and/or total urgent HF visits

6.13.5.3. Continuous Measures

Analyses for the following continuous measures (but not limited to these) will be summarized by treatment and visit:

- weight, BMI, and waist circumference
- FSG and HbA1c
- UACR and eGFR (CKD-EPI Creatinine-Cystatin Equation 2021, CKD-EPI Creatinine Equation 2021, and CKD-EPI Cystatin C equation 2012), and
- blood lipids: total cholesterol, HDL cholesterol, non-HDL, LDL cholesterol, and triglycerides.
- rate of change in eGFR (CKD-EPI Creatinine-Cystatin Equation 2021, CKD-EPI Creatinine Equation 2021 and CKD-EPI Cystatin C equation 2012) per year in patients with high or very-high risk CKD

For BMI, waist circumference, actual value, and change from baseline will be provided. For weight, UACR, and blood lipids, percent change from baseline will also be provided in addition to actual value and change from baseline. For UACR and blood lipids, log transformation values will be used in the analysis model. The results will be provided in the reports with the scale of back transforming to the original, related scale.

The analysis will be conducted based on a mixed-effects model for repeated measures (MMRM) model. Restricted maximum likelihood (REML) will be used to obtain model parameter estimates, and a Kenward-Roger option will be used to estimate denominator degrees of freedom. The response variable of the MMRM model will include treatment, visit, treatment by visit interaction, country/pooled country, and SGLT-2i use (Yes/No) at baseline as fixed effects and baseline value as a covariate. An unstructured covariance structure will be used to model the within-patient errors. If this model fails to converge, the following covariance structures will be tested in order: heterogeneous Toeplitz, heterogeneous first order autoregressive, heterogeneous compound symmetry, Toeplitz, first-order autoregressive, and compound symmetry. The first

covariance structure that converges will be used. Line plots for LS mean with SE will be provided. Rate of change in eGFR (CKD-EPI Cystatin C equation 2012) per year will be analyzed using the same approach as eGFR (CKD-EPI Creatinine Equation 2021) described in Section 6.13.3.

Rate of change in eGFR per year (eGFR slope) will be compared using a mixed effect model. Restricted maximum likelihood (REML) will be used to obtain slope parameter estimates, and a Kenward-Roger option will be used to estimate denominator degrees of freedom. The model for absolute eGFR values as response value will include treatment, time, treatment by time interaction, country/pooled country, and SGLT-2i use (Yes/No) at baseline. Where time is a continuous variable of years from randomization. Intercept and time will be treated as random effect between patients. An unstructured covariance structure will be used to model the within-patient errors. Difference in rate of change in eGFR per year between treatments will be represented as the coefficient estimate of treatment by time interaction.

On-Treatment Analysis

The analyses described above will be repeated using data during the first on-treatment observation period.

6.13.5.4. Categorical Measures

The following categorical measure (but not limited to these) will be summarized by treatment and visit, where the main time point of interest is 36 months:

- proportion of patients with ≥5, 10, 15 and 20% weight loss from baseline.
- proportion of patients with <5.7% and ≤6.5% HbA1c.

Counts and proportions of patients who achieve the target weight loss from screening will be provided. Data from all planned visits to measure weight through the follow-up period will be used to analyze proportion(s) of patients with ≥10% weight loss from screening using a longitudinal logistic regression with repeated measurements, conducted by a generalized, linear mixed model including treatment, country/pooled country, SGLT-2i use (Yes/No) at baseline, baseline value, visit, and treatment-by-visit interaction. An unstructured covariance structure will be used to model the within-patient errors. If this model fails to converge, the following covariance structures will be tested in order: heterogeneous Toeplitz, heterogeneous first-order autoregressive, heterogeneous compound symmetry, Toeplitz, first-order autoregressive, and compound symmetry. The first covariance structure that converges will be used.

On-Treatment Analysis

The analyses described above will be repeated using data during the first on-treatment observation period.

6.13.6. Exploratory Analyses

6.13.6.1. Indirect Comparison versus Placebo for Primary and Key Secondary MACE Endpoints

An indirect treatment effect of tirzepatide compared with placebo will be estimated for the primary and CV related key secondary endpoints, via following steps.

1. "Indirect Treatment Comparison Target Population" will include ITT patients from REWIND who would have been eligible to enroll in SURPASS-CVOT represented by a baseline HbA1c ≥7.0% and a history of at least one of the following: MI, >50% stenosis, coronary revascularization, ischemic stroke, carotid artery revascularization, ankle-brachial index <0.9, peripheral revascularization (iliac or femoral artery), or amputation, and all mITT patients from SURPASS-CVOT.
2. The propensity score, defined as the probability of being enrolled in SURPASS-CVOT relative to REWIND, will be estimated based on covariates including age, sex, baseline BMI, history of coronary revascularization, history of heart failure, history of MI, history of stroke, T2D duration, baseline HbA1c, baseline SBP, baseline eGFR (CKD-EPI Creatinine Equation 2021), logarithm of baseline UACR, baseline non-HDL and smoking habit (current or non-current). The propensity score estimation will also account for missing covariates guided by the multiple imputation missingness pattern (MIMP) approach (Qu and Lipkovich 2009), generating multiple (e.g. 100) complete datasets.
3. Steps 3-1, 3-2 and 3-3 below will be repeated for each complete dataset. In steps 3-1 and 3-2, the stabilized inverse-probability weight calculated based on the estimated propensity score from step 2 where weights > 10 will be replaced by 10 (Stürmer et. Al. 2014) is used to estimate the weighted hazard ratio, and the robust sandwich approach is used to obtain variance of weighted estimates of hazard ratio.

   3-1 In SURPASS-CVOT mITT patients, estimate the hazard ratio of tirzepatide versus dulaglutide based on a Cox proportional hazard model stratified by SGLT-2i use at baseline; HR (SURPASS-CVOT tirzepatide vs SURPASS-CVOT dulaglutide).

   3-2 In REWIND target population, estimate the hazard ratio of dulaglutide versus placebo based on a Cox proportional hazard model; HR (REWIND dulaglutide vs REWIND placebo).

   3-3 Derive the indirect treatment effect of tirzepatide versus placebo as HR (SURPASS-CVOT tirzepatide vs SURPASS-CVOT dulaglutide) multiplied by HR (REWIND dulaglutide vs REWIND placebo).
4. The summarized weighted estimates from step 3-3 (ie, log[HR]s and the corresponding standard errors) will be obtained using Rubin's formula, and then the scale of back transforming to the HR, its 95% CI and p-value will be presented.

6.13.6.2. eGFR change over time

The change in eGFR from baseline to each postbaseline time points will be analyzed in the similar manner as for the key secondary analysis for the change from baseline to 36 months in patients with high or very-high risk chronic kidney disease (CKD) (KDIGO guideline 2025 definition) at baseline. The missing data will be handled in the same way as the key secondary analysis for the change in eGFR from baseline to 36 months. For time points after 36 months, missing value due to lack of opportunity to complete the study visit due to study end (i.e., late enrollment) will be considered Scenario 5B in the imputation.

Additionally, missing data due to any premature study discontinuation (not due to study closure) may be imputed using retrieved dropouts as an additional sensitivity analysis for the change in eGFR from baseline to all postbaseline time points.

6.14. Safety Analyses

Unless specified otherwise, safety summaries and analyses will be carried out using the Safety dataset (see Table GPGN.6.1). All events that occur between the date of the first dose of study drug to the date of the patient's end of follow-up, regardless of the adherence to study drug or initiation of new antihyperglycemic therapy, will be included.

6.14.1. Adverse Events

A TEAE is defined as an event that first occurred or worsened in severity after baseline. The Medical Dictionary for Regulatory Activities (MedDRA) Lowest Level Term (LLT) will be used in the treatment-emergent derivation. The maximum severity for each LLT during the screening period, including ongoing medical history, will be used as baseline severity. Events with a missing severity during the baseline period will be treated as "mild" in severity for determining treatment-emergence. Events with a missing severity during the postbaseline period will be treated as "severe," and treatment-emergence will be determined by comparing to baseline severity.

Statistical comparisons will be applied at both the System Organ Class (SOC) and Preferred Term (PT) levels. For events that are sex-specific, the denominator and computation of the percentage will include only patients from the given sex.

Overviews of the numbers and percentages of patients who experience TEAEs, SAEs, or discontinue from study treatment due to an AE will be summarized by treatment.

Patient narratives will be provided for all patients who experience any of the following "notable" events:

- deaths
- SAEs recorded on AE CRF,
- liver injuries, and
- gallbladder diseases.

6.14.1.1. Treatment-Emergent Adverse Events

The number and percentages of patients with TEAEs will be summarized by treatment using MedDRA PT nested within SOC, and by decreasing frequency across PTs separately. Common (common TEAEs occur in ≥5% of patients before rounding) will be summarized by MedDRA PT, and TEAEs by maximum severity will be summarized by treatment using MedDRA PT nested within SOC.

6.14.1.2. Deaths

The number and percentages of patients with TEAEs resulting in deaths will be summarized by treatment using MedDRA PT nested within SOC.

A listing of all deaths will be provided. The listing will include treatment, patient ID including the site number, date of death, age at the time of enrollment, sex, time from first dose of study drug to death, time from last dose of study drug to death (if patient had discontinued study drug), investigator-reported cause of death, and CEC -confirmed cause of death.

6.14.1.3. Other Serious Adverse Events

The number and percentage of patients who experience an investigator reported SAE will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC.

A listing of SAEs (except ones related to primary, secondary, and additional efficacy endpoints not considered as possibly related to study drug, the drug delivery system, or study procedure) will be provided. The listing will include treatment, patient ID including the site number, date of event, age at the time of enrollment, gender, race, MedDRA SOC and PT, severity, outcome, relationship to study drug, time from first dose of study drug to the event, and time from most recent dose of study drug to the event (if the patient discontinued study drug prior to the event).

6.14.1.4. Discontinuation from Study Due to Adverse Event

The number and percentage of patients who prematurely discontinue the study due to an AE will be summarized by treatment using the MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC.

6.14.1.5. Discontinuation from Study Drug Due to Adverse Event

The number and percentage of patients who prematurely discontinue study drug due to an AE will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC.

The HR and its 95% CI and p-value of time to study drug discontinuation due to an AE will be provided in the same manner as the primary endpoint analysis model. The estimate of the cumulative-event curve over time will be obtained using the KM method. Counts and proportions of patients who experience an event will be calculated.

6.14.1.6. Treatment of Overdose

A listing of patients reporting AEs related to overdosing of tirzepatide and dulaglutide will be provided.

6.14.2. Special Safety Topics

This section includes analyses for safety topics of interest. In general, safety topics of interest will be identified by one or more standardized Medical Dictionary for Regulatory Activities (MedDRA) query(ies) (SMQs), by a Lilly defined MedDRA Preferred Terms (PT) listing based upon the review of the most current MedDRA Version, or by relevant laboratory changes.

The planned analyses for safety topics of interest are provided in the following sections. Detailed searching criteria can be found in compound-level safety standards.

6.14.2.1. Clinically Significant and Severe Hypoglycemia

Definitions of different categories of hypoglycemic events described below will be assessed.

Clinically Significant Hypoglycemia (Level 2)

- Documented, symptomatic hypoglycemia is defined as any time a patient feels that he or she is experiencing symptoms and/or signs associated with hypoglycemia and has a blood glucose (BG) level <54 mg/dL (<3.0 mmol/L).
- Documented, asymptomatic hypoglycemia is defined as any event not accompanied by typical symptoms of hypoglycemia but with a measured BG <54 mg/dL (<3.0 mmol/L).
- Documented, unspecified hypoglycemia is defined as any event with no information about symptoms of hypoglycemia available but with a measured BG <54 mg/dL (<3.0 mmol/L).

Severe Hypoglycemia (Level 3)

- Severe hypoglycemia is defined as an episode with severe cognitive impairment requiring the assistance of another person to actively administer carbohydrate, glucagon, or other resuscitative actions.

Nocturnal Hypoglycemia

- Nocturnal hypoglycemia is defined as any hypoglycemic event that occurs between bedtime and waking.

To avoid duplicate reporting, all consecutive hypoglycemic events occurring within a 1-hour period may be considered a single hypoglycemic event.

Severe Hypoglycemia with Significant Clinical Characteristics

- Severe hypoglycemia with significant clinical characteristics is defined as a severe hypoglycemia with a measured BG <54 mg/dL (<3.0 mmol/L), and
- Is defined as an episode with severe cognitive impairment requiring the assistance of another person to actively administer carbohydrate, glucagon, or other resuscitative actions, and/or hypoglycemia with a measured BG <54 mg/dL (<3.0 mmol/L).

The incidence of hypoglycemic events between at various intervals and overall will be analyzed using a logistic regression with treatment and SGLT-2i use at baseline as fixed effects.

The rate of hypoglycemic episodes per patient-year, by intervals specified above, and overall will be analyzed using a generalized linear mixed-effects model assuming the number of hypoglycemic episodes follows a negative binomial distribution, with treatment, SGLT-2i use at baseline as fixed effects. The logarithm of days during the corresponding period between visits divided by 365.25 days will be adjusted as an offset.

The incidence (percent of participants experiencing ≥1 episode) and rate (episodes/participant/year) of level 2 or level 3, level 3 hypoglycemia and nocturnal level 2 or level 3 hypoglycemia, will be analyzed for overall population and by status of baseline insulin and/or SU use.

6.14.2.2. Pancreatitis

Summaries of adjudicated-confirmed and investigator-reported pancreatic events along with the event rate adjusted by follow-up duration will be provided by treatment.

A summary of adjudicated pancreatic events will be provided by treatment. The total person-years of follow-up and the incidence rate per 100 person-years of follow-up will be provided. Person-years of follow-up is the sum of time to the onset of the event since randomization for participants with the event, and time to the end of follow-up since randomization for participants without the event. The incidence rate per 100 person-years of follow-up is defined by dividing the number of participants who developed the event during the follow-up period by total person-years of follow-up multiplied by 100.

6.14.2.2.1. Pancreatic Enzyme Assessment

Observed pancreatic enzyme data (p-amylase and lipase) will be summarized by treatment and nominal visit. The number and proportion of patients with maximum postbaseline pancreatic enzyme values meeting the following thresholds will be provided (≤1X ULN, >1 to ≤3X ULN, >3 to ≤5X ULN, >5 to ≤10X ULN, and >10X ULN) by baseline maximum pancreatic enzyme value (≤1X ULN, >1 to ≤3X ULN, and >3X ULN).

An MMRM analysis will be conducted to analyze the actual value, absolute, and percent change from baseline using log transformation values in the same manner described in Section 6.13.5.4.

6.14.2.2.2. Pancreatic Malignancies

Pancreatic malignancies will be considered AESI. Treatment-emergent pancreatic malignancies will be identified using the below predefined MedDRA HLT and PT. Summary by treatment and PT will be provided.

Search	Search Term	Code
HLT	Pancreatic neoplasms malignant (excl islet cell and carcinoid)	10033633
PT	Pancreatic neoplasm	10061902

6.14.2.3. Thyroid Malignancies and C-Cell Hyperplasia

Treatment-emergent thyroid malignancies and C-cell hyperplasia will be identified using predefined MedDRA terms. A summary by treatment and PT will be provided.

6.14.2.3.1. Calcitonin

Observed calcitonin data will be summarized by baseline eGFR (CKD-EPI Creatinine-Cystatin Equation 2021) (<, ≥60 mL/min/1.73 m^2), treatment, and visit.

An MMRM model will be fitted to each eGFR category group (<, ≥60 mL/min/1.73 m^2) to analyze the actual value, absolute, and percent change from baseline using log-transformed values in the same manner described in Section 6.13.5.4.

The number and proportion of patients with maximum postbaseline calcitonin value meeting the following thresholds will be provided by baseline eGFR (<, ≥60 mL/min/1.73 m^2), treatment and baseline calcitonin value (≤20 ng/L, >20 ng/L to ≤35 ng/L, >35 ng/L), and postbaseline calcitonin categories (≤20 ng/L, >20 ng/L to ≤35 ng/L, >35 ng/L to ≤50 ng/L, >50 ng/L to ≤100 ng/L, >100 ng/L).

6.14.2.4. Cardiac Arrhythmias and Conduction Disorders

The AE database will be searched using predefined MedDRA terms to identify events consistent with cardiac arrhythmias and conduction disorders. Incidence of the resulting TEAEs will be summarized by treatment and PT.

6.14.2.5. Hypersensitivity Events

Hypersensitivity reactions will be summarized by treatment.

All potential hypersensitivity reactions will be reported by PT within Standardized MedDRA Queries (SMQs narrow scope). Preferred Terms will be used for the summary within each category in decreasing order of incidence.

6.14.2.6. Diabetic Retinopathy Complications

Results of the baseline dilated fundoscopic exam will be summarized by treatment. A summary of TEAEs suspected of worsening retinopathy by treatment and PT and a summary of change from baseline in the dilated fundoscopic exam will be summarized. Any TEAEs suspected of worsening retinopathy that are confirmed by corresponding changes in the dilated fundoscopic exam from baseline will also be summarized by treatment and PT. Detailed analyses of diabetic retinopathy assessments are conducted for the participants who enrolled in diabetic retinopathy substudy as described in Section 6.14.5.

In addition, the first occurrence of the composite diabetic retinopathy endpoint consisting of 1) retinal photocoagulation, 2) vitreous hemorrhage, 3) vitrectomy, 4) treatment with intravitreal agents, or 5) onset of diabetes-related significant visual loss, the first occurrence of each component will be analyzed as follows.

- The HR, with its 95% CI and p-value, will be provided in the same manner as the primary endpoint analysis model.
- The estimate of the cumulative event curve over time will be obtained using the KM method.
- Counts and proportions of patients who experience an event will be calculated.

6.14.2.7. Hepatobiliary Disorders

6.14.2.7.1. Acute Gallbladder Disease

Summaries of all events of treatment-emergent biliary colic, cholecystitis, or other suspected events related to gallbladder disease will be generated by PT with decreasing frequency by treatment.

6.14.2.7.2. Liver Enzymes

Analyses for laboratory analyte measurements are described in Section 6.14.4. This section describes additional analyses of liver enzymes. The number and percentage of patients at any time during the study with the following thresholds will be provided by treatment group.

- ALT: The number and percentage of participants with a measurement greater than or equal to 1 time (1X), 3 times (3X), 5 times (5X), 8 times (8X), 10 times (10X), and 20 times (20X) the performing lab upper limit of normal (ULN) during the postbaseline period for all participants with a postbaseline value.
- AST: The number and percentage of participants with a measurement greater than or equal to 1 time (1X), 3 times (3X), 5 times (5X), 8 times (8X), 10 times (10X), and 20 times (20X) the performing lab upper limit of normal (ULN) during the postbaseline period for all participants with a postbaseline value.
- ALP: The number and percentage of participants with a measurement greater than or equal to 2 times (2X), and 3 times (3X) the performing lab ULN during the postbaseline period will be summarized for all participants with a postbaseline.
- TBL: The number and percentage of participants with a measurement greater than or equal to 2 times (2X), 5 times (5X), and 8 times (8X) the performing lab ULN during the postbaseline period will be summarized for all participants with a postbaseline value.
- DBL: The number and percentage of participants with a measurement greater than or equal to 2 times (2X) and 5 times (5X) the performing lab ULN during the postbaseline period will be summarized for all participants with a postbaseline value.
- Hepatocellular Drug Induced Liver Injury Screening Plot (TBL vs ALT or AST)
- Hepatocellular Drug Induced Liver Injury Screening Table
- Cholestatic Drug Induced Liver Injury Screening Plot (TBL vs ALP)
- Cholestatic Drug Induced Liver Injury Screening Table

For the postbaseline maximum value, all planned and unplanned measurements will be included. When/if multiple values are available (ie, unplanned measurement) prior to randomization, the maximum value will be used as baseline.

6.14.2.7.3. Hepatic Treatment-Emergent Adverse Events

The percentages of patients with treatment-emergent, potentially drug-related hepatic disorders will be summarized by treatment using MedDRA PTs within SMQs.

6.14.2.8. Gastrointestinal Safety

Summaries of all events of treatment-emergent AEs in GI disorders SOC, including nausea, vomiting, diarrhea and constipation will be provided by treatment and PT with decreasing frequency.

The time courses of prevalence and incidence (newly occurring episodes) of GI AEs (nausea, vomiting, diarrhea, constipation and combined nausea, diarrhea, and vomiting) will be plotted by treatment group and severity.

The maximum severity of treatment-emergent nausea, vomiting, diarrhea, constipation and combined nausea, diarrhea, and vomiting through the end of the study will be summarized by treatment.

6.14.2.9. Acute Renal Failure or Exacerbation of Chronic Renal Failure

Laboratory measures related to renal safety including eGFR (both eGFRs based on creatinine and cystatin-C) will be analyzed as specified for laboratory measurements in Section 6.14.4. eGFR based on creatinine will be derived using the CKD-EPI Creatinine-Cystatin Equation 2021, whereas the central laboratory reported values based on a separate equation (*Ann Intern Med* 2009; 150:604–612) are used to review an exclusion criterion ([18] <15 mL/minute/1.73 m^2).

Two shift tables examining renal function will be created with the following thresholds by treatment group.

- Minimum baseline to minimum postbaseline through the study for eGFR - <30, ≥30 to <45, ≥45 to <60, ≥60 to <90, ≥90 mL/min/1.73m2.
- Maximum baseline to maximum postbaseline through the study for UACR - <30, ≥30 to ≤300, >300 mg/g (respectively, these represent normal, microalbuminuria, and macroalbuminuria).

When/if multiple values are available (ie, unplanned measurement) prior to randomization, the minimum value for eGFR and maximum value for UACR will be used as baseline.

To examine AEs for any suggestion of decrease in renal function, a summary of acute renal events searched by acute renal failure and chronic kidney disease MedDRA SMQs will be presented.

Analysis for the following composite diabetic nephropathy-endpoints for overall and patients with high or very high-risk CKD will be analyzed in the same manner specified as an event related-efficacy measure in Section 6.13.5.3 for mITT datasets because these endpoints will be served as part of efficacy topic.

- composite renal endpoint 1: 1) persistent macroalbuminuria, 2) persistent doubling of serum creatinine level and creatinine clearance (per eGFR<45 mL/min/1.73m2), 3) need for continuous renal-replacement therapy, or 4) death due to renal disease,
- composite renal endpoint 2: a) onset of persistent ≥40% reduction in eGFR, b) ESRD, or c) death due to renal disease

- composite renal endpoint 3: i) onset of renal failure, defined as initiation of chronic renal replacement therapy (dialysis or kidney transplantation) or persistent eGFR<15 mL/minute/1.73 m^2, ii) death due to renal disease, iii) CV death, or iv) onset of persistent ≥50% reduction in eGFR

Onset of "persistent" is determined as the date when a patient met criteria at or after the 1-year visit and continued to meet criteria at all subsequent visits. If a patient did not meet criteria at the measurement prior to the final measurement, but met criteria at the final measurement, the date of onset of "persistent" is the date of the last measurement.

6.14.2.10. Amputation/Peripheral Revascularization

The AE database will be searched using MedDRA PT to identify events consistent with amputation, peripheral revascularization, and other PTs containing the word of ”amputation”. The incidence of the resulting TEAEs will be summarized by treatment and PT.

6.14.3. Vital Signs and Other Physical Findings

Descriptive summaries by treatment and by nominal visit will be provided for baseline and postbaseline values as well as change from baseline values. If 2 records are taken at the same visit, they will be averaged prior to being used for data summaries and analyses.

An MMRM model using REML will be used to fit the changes from baseline in vital signs at all scheduled postbaseline visits. The model will include treatment, visit, treatment-by-visit interaction, country/pooled country, and SGLT-2i use at baseline (Yes/No) as fixed effects and baseline value of the dependent variable as a covariate.

Counts and percentages of patients with treatment-emergent abnormal sitting systolic BP, sitting diastolic BP, and pulse will be presented by treatment group and nominal visit. The criteria for identifying patients with treatment-emergent vital sign abnormalities are stated in Table GPGN.6.5.

Table GPGN.6.5. Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurements

Parameter	Low	High
Systolic BP (mm Hg) (Supine or sitting – forearm at heart level)	≤90 and decrease from baseline ≥20	≥129 and increase from baseline ≥20 ≥140 and increase from baseline ≥20
Diastolic BP (mm Hg) (Supine or sitting – forearm at heart level)	≤50 and decrease from baseline ≥10	≥90 and increase from baseline ≥10
Pulse (bpm) (Supine or sitting)	<50 and decrease from baseline ≥15	>100 and increase from baseline ≥15

Abbreviation: BP = blood pressure.

6.14.4. Clinical Laboratory Evaluation

All laboratory data will be reported both in the International System of Units and Conventional units.

An MMRM analysis will be used to analyze the actual value, absolute, and percent change from baseline for selected laboratory measures with or without log-transformed (postbaseline measure/baseline measure) response variables, and with treatment, country/pooled country, SGLT-2i use, visit, and treatment-by-visit interaction as fixed effects and baseline value (log transformed for log-transformed response variables) as a covariate. For measures analyzed using log-transformed values, the results will be presented in the reports with the scale back transforming to the original, related scale.

Shift tables including the number and percentage of patients within each baseline category versus each maximum postbaseline category for selected measurements will be provided by treatment group. The proportion of patients shifted will be compared between treatments. Shift table from normal/high to low or from normal/low to high will be summarized for lab parameters.

6.14.5. Diabetic Retinopathy in Patients Enrolled in Substudy

This section is applicable to the participants who are enrolled in the Diabetic Retinopathy Substudy.

6.14.5.1. General Consideration

In time-to-event analyses for endpoints assessed by standard retinal fundus photographs and visual acuity checks, the event date will be defined as the date of the first assessment when the event is confirmed if the participant experiences the event; if the participant does not experience the event, the censoring date will be the date of participant's last assessment date of endpoint (that is, standard retinal fundus photographs, or visual acuity checks). For all other endpoints, the censoring date will be participant's end of follow-up date as specified in Section 6.2.

For sustained endpoint analyses, the onset of "sustained" is defined as the date of assessment when a participant first met the endpoint criteria and continues to meet the criteria at all subsequent visits, including the final assessment prior to the initiation of diabetic retinopathy (DR) treatment during the follow-up period. Participant who received DR treatment at any time during the follow-up period, will be considered to have experienced a sustained endpoint event, with the onset date defined as the date of DR treatment initiation. If a participant initially meets the endpoint criteria but later did not meet the criteria at a subsequent visit or meets thecriteria at the final assessment, without prior DR treatment, the onset date of "sustained" is defined as the date of the final assessment.

DR treatment is defined as treatment with focal laser photocoagulation, panretinal laser photocoagulation, or intravitreal injection. The assessment report, provided by third-party graders based on standard retinal fundus photographs, includes the Early Treatment Diabetic Retinopathy Study (ETDRS) scale and Diabetic Retinopathy Severity Scale (DRSS) level for each eye. ETDRS scale per person will be derived based on DRSS level for either or both eyes according to ACCORD Eye Study (The ACCORD Study Group & ACCORD Eye Study Group, 2010) (see Section 8.1, Appendix 1) and used for analyses, instead of grader's scale for each eye, unless otherwise stated.

6.14.5.2. Patient Characteristics

Selected demographic, medical history, concomitant illness, and baseline diabetic retinopathy assessments will be summarized by treatment group.

6.14.5.3. Primary Endpoint Analysis

The primary objective of the substudy is to compare the effect of tirzepatide dose up to 15 mg QW with dulaglutide 1.5 mg QW on DR progression from baseline to 36 months visit for at least 2 grades worsening of diabetic retinopathy (DR) per person on Early Treatment Diabetic Retinopathy Study (ETDRS) Diabetic Retinopathy Severity Scale (DRSS) - (ETDRS2+). The analyses for this endpoint will be conducted using the safety dataset.

This endpoint will be analyzed using a logistic regression model with treatment, baseline SGLT-2i use (Yes/No), baseline HbA1c, and duration of T2DM as fixed effects, and baseline ETDRS scale as a covariate. Missing ETDRS scale data will be imputed as specified in 6.14.5.3.1. The unconditional treatment group effect will be assessed by odds ratio on the delta-method using the formula provided (Ye et al. 2023, FDA 2023). The final inference will be derived using Rubin's Rule by combining estimates from multiple imputed datasets. The risk difference, odds ratio, p-value, and 95% CI will be presented.

If the upper limit of the 95% CI of odds ratio (tirzepatide versus dulaglutide) is less than 1.8, the substudy will conclude that tirzepatide is not associated with clinically meaningful increased risk for worsening DR status compared to dulaglutide.

This primary model will be repeated using the missing data imputation approach as described in Section 6.14.5.3.1, for each timepoint, i.e., week 24, months 12, 18 and 24.

6.14.5.3.1. Methods for Missing Data Imputation

The missing ETDRS scale will be imputed based on the reason for missingness.

1. Participants who did not receive diabetic retinopathy treatment during the follow-up period but have missing ETDRS scale values will have these ETDRS scale imputed using multiple imputation under the assumption of missing at random (MAR)
2. At the visit of interest, transform the observed and imputed continuous value to the binary value.
3. Participants who received diabetic retinopathy treatment during the follow-up period will be considered to have experienced an event on the date of diabetic retinopathy treatment and assume to continue thereafter.

6.14.5.3.2. Sensitivity Analysis of Primary Endpoint

For the primary endpoint, a sensitivity analysis will be performed using a different missing data imputation from the one described in Step 3 of Section 6.14.5.3.1. For participants who received diabetic retinopathy treatment, ETDRS scale collected after the initiation of treatment will be replaced with the maximum observed value from baseline to Month 36.

6.14.5.4. Key Secondary Endpoint Analysis

The analyses for the following key secondary endpoints will be conducted using the safety dataset.

Relative to the Endpoint Measure	Analysis Conducted	Additional Information
Time to first onset of ETDRS2+ per person	Time-to-event: Cox Regression; The HR with its 95% CI and p-value will be provided	Refer Section 6.14.5.3.1 for missing data imputation
Time to first onset of sustained worsening of visual acuity of at least 3 lines (15 letters) on either eye		A worsening of visual acuity by 15 letters corresponds to a 0.3 increase in logMAR[a] from baseline.

[a] Snellen fraction (decimal VA) will be converted to logMAR, using the formula logMAR = $-\log_{10}$ * decimal VA.

A Cox proportional hazard model stratified by SGLT-2i use (Yes/No) at baseline with treatment, baseline HbA1c and duration of T2DM as fixed effects, and baseline value as a covariate will be used for analyses.

6.14.5.5. Exploratory Endpoint Analysis

The exploratory objective is to assess the safety of tirzepatide dose up to 15 mg QW when compared with dulaglutide 1.5 mg QW on DR. The analysis for the following endpoints will be conducted using the safety dataset.

Relative to the Endpoint Measure	Analysis Conducted	Additional Information
ETDRS3+ (defined as at least 3 grades worsening of DR per person on ETDRS scale) from baseline to 36 months visit	Binary: Logistic regression model	Refer to Section 6.14.5.3 for analysis approach detail
ETDRS2+ at 36 months or photocoagulation, intravitreal injections (anti-VEGF, steroid treatment, or both), or vitrectomy or Vitreous hemorrhage in response to ETDRS2+ anytime from baseline		
Time to first onset of ETDRS3+	Time-to-event: Cox Regression; The HR with its 95% CI and p-value will be provided	Refer Section 6.14.5.3.1 for missing data imputation
Time to first onset of sustained ETDRS2+ per person		
Time to first onset of sustained ETDRS3+ per person		
Time to first onset of new diagnosis of proliferative diabetic retinopathy, based on grader's report		
Time to first onset new diagnosis of macular edema, based on Dilated Fundoscopic Examination		
Time to first onset of treatment by photocoagulation, intravitreal injections (anti-VEGF, steroid treatment, both), or vitrectomy or Vitreous hemorrhage		
Time to first onset of treatment by photocoagulation, intravitreal injections (anti-VEGF, steroid treatment, or both), or vitrectomy Vitreous hemorrhage followed by ETDRS2+		
Time to first occurrence of treatment for at least one of the following: focal/grid laser photocoagulation pan-retinal laser photocoagulation intravitreal injection with anti-VEGF intravitreal injection with steroid vitrectomy vitreous hemorrhage		
Mean change of ETDRS scale	Continuous: MMRM	Refer to Section 6.13.5.3 for analysis approach detail
Mean change of ETDRS scale by baseline ETDRS (no retinopathy, any retinopathy)		
Change in visual acuity in the worse- or better-seeing eye as measured in the number of letters [a] using best correction		
Change in HbA1c from baseline to Month 36		
Change in systolic and diastolic blood pressure from baseline to Month 36		
Change in Lipids: total-cholesterol, high density lipoprotein (HDL)-cholesterol, low density lipoprotein (LDL)-cholesterol and triglycerides from baseline to Month 36		

[a] For analysis purpose, number of letters will be derived using the formula: number of letters = 85−(50*LogMAR) as described in Elliott (2016).

For the assessment of binary and time to event, the model with treatment, baseline SGLT-2i use (Yes/No), baseline HbA1c, and duration of T2DM as fixed effects, and baseline value as a covariate will be used to examine the treatment difference. For continuous measures, country/pooled country will also be added as a fixed effect.

In addition to above endpoints, following parameters will be summarized by treatment group and by visit.

- Grader reported diabetic retinopathy severity level per eye
- Derived ETDRS scale
- Grader reported diabetes retinopathy status per eye
- Macular edema status per eye based on fundoscopic exam
- Shift between no retinopathy, mild-, moderate-, severe non-proliferative retinopathy or proliferative retinopathy between baseline and each postbaseline timepoint for each eye, based on grader's report
- Presence of at least 2 grade ETDRS DRSS progression and improvement versus baseline anytime post baseline
- Presence of at least 3 grade ETDRS DRSS progression and improvement versus baseline anytime post baseline

Further analyses for investigating correlations between DR progression and biomarkers may be conducted.

6.15. Health Outcomes/Quality-of-Life Analyses

6.15.1. 5-Level European Quality of Life Questionnaire

Each item will be summarized descriptively by treatment at each scheduled visit at which the EQ-5D-5L is administered. The changes from baseline in the index and visual analog scale scores for visits of 24 weeks, 24 months, and final visit (only if mappable measure to scheduled visit) will be analyzed using an MMRM model in a similar manner specified in Section 6.13.5.3. And, for 24 months, the scores will be analyzed using analysis of covariance (ANCOVA) as specified in Section 6.4.1.

6.15.2. Ability to Perform Physical Activities of Daily Living

Descriptive summaries by treatment at each scheduled visit at which the APPADL is administered will be presented for each item. The changes from baseline in the transformed scores for visits of 24 weeks, 24 months, and final visit (only if mappable measure to scheduled visit) will be analyzed using an MMRM model in a similar manner specified in Section 6.13.5.3. And, for 24 months, the scores will be analyzed using analysis of covariance (ANCOVA) as specified in Section 6.4.1.

6.15.3. Impact of Weight on Self-Perceptions Questionnaire

Descriptive summaries by treatment at each scheduled visit at which the IW-SP is administered will be presented for each item. The changes from baseline in transformed scores for visits of 24 weeks, 24 months, and final visit (only if mappable measure to scheduled visit) will be analyzed using an MMRM model in a similar manner specified in Section 6.13.5.3. And, for

24 months, the scores will be analyzed using analysis of covariance (ANCOVA) as specified in Section 6.4.1.

6.16. Subgroup Analyses

6.16.1. Primary Efficacy Endpoint

Subgroup analyses relative to the primary efficacy endpoint will be analyzed.

The HR and its 95% CI will be provided using a Cox proportional hazards model stratified by SGLT-2i use at baseline with subgroup, treatment group, and the interaction between the levels of the subgroup and treatment group as categorical variables. When analyzing SGLT-2i use as a subgroup, the model will not be stratified by SGLT-2i use.

The estimate of the cumulative event curve over time will be plotted by treatment group for each subgroup level using the KM method. Counts and proportions of patients who experience an event will be calculated. A Forest plot of the HR and its 95% CI will be provided along with informative values such as number of events, number of patients (percentage), and p-value for the interaction of treatment by subgroup level (may be modified appropriately).

Planned subgroups will include but are not limited to the following.

Demographics

- sex
- age (<65 years, ≥65 years)
- region of enrollment ([1] North America, South America, Europe, and Asia-Pacific, [2] Ukraine, Russia, and the rest of the world)
    - North America – Canada and the US
    - South America - Argentina, Brazil, and Mexico
    - Europe - Austria, Belgium, Czech Republic, France, Germany, Greece, Hungary, Israel, Italy, Netherlands, Poland, Romania, Russia, Slovakia, Spain, Sweden, Turkey, Ukraine, and the United Kingdom
    - Asia-Pacific - Australia, China, India, Japan, South Korea, and Taiwan
- race
- ethnic group, and
- BMI ([1] <27 kg/m^2, 27-30 kg/m^2, >30 kg/m^2, [2] <27 kg/m^2, >=27 kg/m^2, [3] <35 kg/m^2, ≥35 kg/m^2).

Medical History and Baseline Antihyperglycemic Medication use

- HbA1c (≤8.5%, >8.5%)
- duration of diabetes (<10 years, ≥10 years)

- history of MI (Yes/No)
- history of stroke (Yes/No)
- history of heart failure (Yes/No)
- eGFR (CKD-EPI Creatinine-Cystatin Equation 2021) (<60, 60-<90, >=90 ml/min/1.73 m^2)
- UACR (<30, ≥30 mg/g)
- baseline SGLT-2i use (Yes/No), and
- SGLT-2i use at any time on and after baseline.

A Forest plot of the HR and its 95% CI by country will be provided in ascending order of HR. A funnel plot of HR versus total number of events by country may also be provided.

6.16.2. Heart Failure Related Events

Following analyses will be performed by patients with history of heart failure at baseline (yes/no).

- time to first occurrence of CV death or heart failure event requiring hospitalization and/or urgent HF visit
- time to first occurrence of heart failure event requiring hospitalization and/or urgent HF visit
- cumulative number of CV deaths, total (first and recurrent) HF events requiring hospitalization, and/or total urgent HF visits

6.16.3. Other Endpoints

BMI subgroup (<27 kg/m^2, >=27 kg/m^2) analyses will also be conducted on death due to any cause, and MACE-4.

A subgroup analysis for each common TEAEs (>=5%) will be performed for patients treated with tirzepatide for selected subgroup variables including but not limited to age, sex, race, ethnicity, baseline BMI, baseline eGFR, and region.

An overview of adverse events by the following age categories (<65 years, 65-74 years, 75-84 years, ≥85 years) will be included to meet expectations from the European Union (EMA 2014).

Other subgroup analyses may be done by country to support local regulatory registrations.

6.17. Interim Analyses and Data Monitoring

An interim analysis will be conducted under the auspices of an IDMC according to the specifications set forth in the protocol.

The IDMC is authorized to evaluate unblinded interim efficacy and safety analyses prior to DBL. Study sites will receive information about interim results ONLY if they need to know for the

safety of their patients. The study team will remain blinded until the final DBL is locked unless the study is prematurely terminated. The first scheduled IDMC review will occur approximately 6 months after the first patient visit. Thereafter, IDMC reviews will be every 6 months throughout the study. Tabular summaries of clinical endpoint- events (reported and adjudicated) including deaths, MIs, and strokes, selected potentially clinically significant laboratory values, AEs, SAEs, discontinuations due to AEs, and other relevant safety information by treatment group will be sent to the IDMC by a statistical analysis center external to and independent from the sponsor.

A single, planned interim efficacy analysis will be performed once approximately 1075 CEC-confirmed MACE-3 endpoints have been accrued to evaluate the need for further continuation of the study. Details in alpha-level considering a formal interim efficacy analysis is provided in Section 6.6. Consideration will be given to stop the trial if tirzepatide is superior to dulaglutide. If not, the trial will continue until at least 1615 patients experience at least 1 component of the composite MACE-3 endpoint

Details of interim stopping rules, membership, operations, and the communication plan is documented in the IDMC charter. Details regarding maintaining the blind during the conduct of this study are specified in the blinding and unblinding plan.

If the main study is terminated early for efficacy, patients participating in retinopathy substudy will be offered to continue receiving study intervention and performing assessments until completing the 3 years follow-up. The analyses specific for measurements in the retinopathy sub-study, described in Section 6.14.5, will be conducted after completing substudy.

7. Unblinding Plan

The unblinding plan is specified in a separate document.

8. Supporting Documentation

8.1. Appendix 1: Description of ETDRS Grading Scale

Table APP.1.1. Derivation of ETDRS Scale for Persons (The ACCORD Study Group & ACCORD Eye Study Group, 2010)

Severity Level	Description	Scale step per person
10/10	No DR	1
20/<20	Microaneurysms only, one eye	2
20/20	Microaneurysms only, both eyes	3
35/<35	Mild NPDR, one eye	4
35/35	Mild NPDR, both eyes	5
43/<43	Moderate NPDR, one eye	6
43/43	Moderate NPDR, both eyes	7
47/<47	Moderately severe NPDR, one eye	8
47/47	Moderately severe NPDR, both eyes	9
53/<53	Severe or very severe NPDR, one eye	10
53/53	Severe or very severe NPDR, both eyes	11
60 or 61/<60	Mild PDR and/or SPC, one eye	12
60 or 61/60 or 61	Mild PDR and/or SPC, both eyes	13
65/<65	Moderate PDR, one eye	14
65/65	Moderate PDR, both eyes	15
71 or 75/<71	High risk PDR, one eye	16
71 or 75/71 or 75	High risk PDR, both eyes	17
81 or 85/<81	Advanced PDR, one eye	18
81 or 85/81 or 85	Advanced PDR, both eyes	19

DR – diabetic retinopathy, NPDR – non proliferative diabetic retinopathy, PDR – proliferative diabetic retinopathy, SPC – scars of photocoagulation.

Source: Table 2 of Section 3 in the Supplementary Appendix for ACCORD Eye Study (The ACCORD Study Group & ACCORD Eye Study Group, 2010).

A change from source table is that severity level of 47/43, which corresponds to the scale step of 9 between the rows 47/<47 and 53/<53 in the source table, was thought to be a typo and has been corrected to 47/47.

In this study, severity levels 12, 14 and 15, which were not included in table 2 of supple. appendix for ACCORD Eye Study, are also available with the definitions of 'Absent' for 10 and 12, and 'Questionable Retinopathy' for 14, 15 and 20; therefore, the scale steps will be derived by replacing severity level 12 with 10, and 14 and 15 with 20. Additionally, severity levels 81 and 85 are also available, and mapped to scale steps 18 and 19.

8.2. Appendix 2: Statistical Analysis for China

Analyses will be performed for the following subpopulation:

- participants enrolled in Mainland China and Taiwan.
- participants enrolled in East Asia (mainland China, Taiwan, Korea and Japan).

The analysis methods for this subpopulation will be similar to those described for the main study SAP of GPGN. If there is not a sufficient number of participants in the subpopulation, summary statistics will be provided without statistical inference.

The analyses to be included will be documented in a separate list of analyses which should include disposition, demographics, and selected efficacy and safety endpoints.

9. References

ACCORD Study Group and ACCORD Eye Study Group. Effects of medical therapies on retinopathy progression in type 2 diabetes. *N Engl J Med*. 2010;363(3):233–244. doi:10.1056/NEJMoa1001288.

Aggarwal R, Harling L, Efthimiou E, et al. The effects of bariatric surgery on cardiac structure and function: a systematic review of cardiac imaging outcomes. *Obes Surg*. 2016;26(5):1030-1040. https://doi.org/10.1007/s11695-015-1866-5

Aminian A, Zajichek A, Aterburn DE, et al. Association of metabolic surgery with major adverse cardiovascular outcomes in patients with type 2 diabetes and obesity. *JAMA*. 2019;322(13):1271-1282. https://doi.org/10.1001/jama.2019.14231

Botros FT, Gerstein HC, Malik R, et al. Dulaglutide and Kidney Function-Related Outcomes in Type 2 Diabetes: A REWIND Post Hoc Analysis. Diabetes Care. 2023 Aug 1;46(8):1524-1530. doi: 10.2337/dc23-0231.

Bretz F, Maurer W, Brannath B, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med*. 2009;28(4):586-604. https://doi.org/10.1002/sim.3495

Elliott, D. B. (2016). The good (logMAR), the bad (Snellen) and the ugly (BCVA, number of letters read) of visual acuity measurement. *Ophthalmic and Physiological Optics*, 36(4), 355–358. https://doi.org/10.1111/opo.12310

[FDA] United States Food and Drug Administration. Diabetes mellitus-evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diebetes; availability: guidance for industry. December 2008. Accessed October 28, 2021. https://www.federalregister.gov/documents/2008/12/19/E8-30086/guidance-for-industry-on-diabetes-mellitus-evaluating-cardiovascular-risk-in-new-antidiabetic

[FDA] United States Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness: guidance for industry. November 2016. Accessed October 28, 2021. https://www.fda.gov/media/78504/download

Gerstein HC, Colhoun HM, Dagenais GR, et al; REWIND Investigators. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. *Lancet*. 2019;394(10193):121-130. https://doi.org/10.1016/S0140-6736(19)31149-3

Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics.* 2000;56(2):554-562. https://doi.org/10.1111/j.0006-341x.2000.00554.x

Hernandez AF, Green JB, Janmohamed S, et al; Harmony Outcomes Committees and Investigators. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. *Lancet*. 2018;392(10157):1519-1529. https://doi.org/10.1016/S0140-6736(18)32261-X

Holman RR, Bethel MA, Mentz RJ, et al; EXSCEL Study Group. Effects of once-weekly exenatide on cardiovascular outcomes in type 2 diabetes. *N Eng J Med*. 2017;377(13):1228-1239. https://doi.org/10.1056/NEJMoa1612917

Husain M, Birkenfeld AL, Donsmark M, et al; PIONEER 6 Investigators. Oral semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2019;381:841-851. https://doi.org/10.1056/NEJMoa1901118

Hwang IK, Shih WJ, De Cani JS. Group sequential designs using a family of type I error probability spending functions. *Stat Med*. 1990;9(12):1439-1445. https://doi.org/10.1002/sim.4780091207

Kidney Disease: Improving Global Outcomes (KDIGO) ADPKD Work Group. KDIGO 2025 Clinical Practice Guideline for the Evaluation, Management, and Treatment of Autosomal Dominant Polycystic Kidney Disease (ADPKD). *Kidney Int*. 2025;107(2S):S1–S239. https://doi.org/10.1016/j.kint.2024.07.009

Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B.* 2000;62:711–730. https://doi.org/10.1111/1467-9868.00259

Marso SP, Bain SC, Consoli A, et al; SUSTAIN-6 Investigators. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2016a;375:1834-1844. https://doi.org/10.1056/NEJMoa1607141

Marso SP, Daniels GH, Brown-Frandsen K, et al; LEADER Steering Committee; LEADER Trial Investigators. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2016b;375(4):311-322. https://doi.org/10.1056/NEJMoa1603827

McGuire DK, D'Alessio D, Nicholls SJ, Nissen SE, Riesmeyer JS, Pavo I, et al; Transitioning to active-controlled trials to evaluate cardiovascular safety and efficacy of medications for type 2 diabetes. *Cardiovascular diabetology.* 2022;21(1):163.

Perkovic V, Tuttle KR, Rossing P, et al; FLOW Trial Committees and Investigators. Effects of Semaglutide on Chronic Kidney Disease in Patients with Type 2 Diabetes. N Engl J Med. 2024 Jul 11;391(2):109-121. doi: 10.1056/NEJMoa2403347. Epub 2024 May 24.

Pfeffer MA, Claggett B, Diaz R, et al; ELIXA Investigators. Lixisenatide in patients with type 2 diabetes and acute coronary syndrome. *N Engl J Med*. 2015;373(23):2247-2257. https://doi.org/10.1056/NEJMoa1509225

Qu Y, Lipkovich I. Propensity score estimation with missing values using a multiple imputation missingness pattern (MIMP) approach. *Stat Med*. 2009;28(9):1402-14. https://doi: 10.1002/sim.3549.

Sattar N, Lee MMY, Kristensen SL, et al. Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of randomised trials. *Lancet Diabetes Endocrinol*. 2021;9: 653–62.

Stürmer T, Wyss R, Glynn RJ, Brookhart MA. Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs. *J Intern Med*. 2014;275(6):570-80. https://doi: 10.1111/joim.12197.

Sundström J, Bruze G, Ottosson J, et al. Weight loss and heart failure: a nationwide study of gastric bypass surgery versus intensive lifestyle treatment. *Circulation*. 2017;135(17):1577-1585. https://doi.org/10.1161/CIRCULATIONAHA.116.025629

Zinman B, Wanner C, Lachin JM, et al; EMPA-REG OUTCOME Investigators. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med*. 2015;373(22):2117-2128.

Signature Page for VV-CLIN-070061 v3.0

Approval	PPD Statistician 24-Jul-2025 17:26:22 GMT+0000

Signature Page for VV-CLIN-070061 v3.0