Statistical Analysis Plan: I8F-MC-GPIH

A Phase 4, Randomized, Open-Label, Active-Controlled Study to Investigate the Efficacy and Safety of Switching from Weekly Dulaglutide to Weekly Tirzepatide in Adults with Type 2 Diabetes

NCT05564039

Approval Date: 01-Dec-2022

Title Page

Protocol Title: A Phase 4, Randomized, Open-Label, Active-Controlled Study to Investigate the Efficacy and Safety of Switching from Weekly Dulaglutide to Weekly Tirzepatide in Adults with Type 2 Diabetes

Protocol Number: I8F-MC-GPIH

Compound Number: tirzepatide (LY3298176)
Short Title: A study to investigate the efficacy and safety of switching from weekly dulaglutide to weekly tirzepatide in adults with type 2 diabetes

Acronym: SURPASS-SWITCH

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana, USA 46285

Regulatory Agency Identifier Number(s)

IND: 128801

EU Trial Number: 2022-500101-41-00

Document ID: VV-CLIN-056526

Table of Contents

Version history

This Statistical Analysis Plan (SAP) for Study I8F-MC-GPIH (GPIH) is based on Protocol GPIH dated 01 July 2022.

Table 1. SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	See date on page 1	Not Applicable	Original version

Abbreviation: SAP = statistical analysis plan.

1. Introduction

This study is designed to investigate the efficacy and safety of switching once-weekly (QW) dulaglutide 0.75 mg or 1.5 mg to QW tirzepatide 15 mg or maximum tolerated dose (MTD), or continuing and escalating to dulaglutide 4.5 mg or MTD in adults with type 2 diabetes (T2D) who are currently on a nonmaximal, stable dose of dulaglutide weekly. Data from this study will show if switching treatment from weekly dulaglutide to tirzepatide rather than intensifying the dulaglutide dose provides additional efficacy in adults with T2D.

This SAP describes the prespecified statistical analyses for Study GPIH. These analyses apply to efficacy and safety data. Changes to the protocol-planned analyses are described in Section 4.9.

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Primary	
To demonstrate that switching from once-weekly dulaglutide to once weekly tirzepatide is superior to continuing and escalating dulaglutide for HbA1c change from baseline to Week 40 in participants with T2D	Change from baseline in HbA1c
Key Secondary	
To demonstrate that switching from once-weekly dulaglutide to once-weekly tirzepatide is superior to continuing and escalating dulaglutide for weight change from baseline to Week 40 in participants with T2D	Change from baseline in weight
Additional Secondary	

To assess the treatment effect of switching from once-weekly dulaglutide to once-weekly tirzepatide compared to continuing and escalating dulaglutide at Week 40 in participants with T2D	• HbA1c <7%, ≤6.5%, and <5.7%
	• Weight loss from baseline of ≥5%, ≥10%, and ≥15%
	• A composite of
	  ○ HbA1c ≤6.5%
	  ○ weight loss ≥10% from baseline
	  ○ no hypoglycemia, defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia

To assess the treatment effect of switching from once-weekly dulaglutide to once-weekly tirzepatide compared to continuing and escalating dulaglutide from baseline to Week 40 in participants with T2D	Change from baseline in • FSG • waist circumference • BMI • Impact of Weight on Quality of Life-Clinical Trials Version (IWQOL-Lite-CT) - Physical Functioning Score
Exploratory	
To assess the treatment effect of switching from once-weekly dulaglutide to once-weekly tirzepatide compared to continuing and escalating dulaglutide from baseline to Week 40 in participants with T2D	Change from baseline in • lipids (total cholesterol, HDL, LDL, VLDL, and TG) • daily average 7-point SMBG profile

Objectives	Endpoints
	• patient-reported outcomes: o IWQOL-Lite-CT: Total, Physical, and Psychosocial Scores o IW-SP o APPADL o EIDTQc o GIEH - 1 week recall o GIEH - since study start recall
Exploratory (not specified in the protocol)	
To demonstrate that switching from once-weekly dulaglutide to once-weekly tirzepatide is superior to continuing and escalating dulaglutide for weight change from baseline to Week 40 in participants with T2D	• Percent change from baseline in body weight Change from baseline in • Proportion of patients who achieved composite endpoint ("HbA1c <5.7% & Weight Loss ≥15% & No-Hypo"), defined as o HbA1c <5.7% o weight loss ≥15%, and o no hypoglycemia, defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia

Abbreviations: APPADL = Ability to Perform Physical Activities of Daily Living; BG = blood glucose; BMI = body mass index; EIDTQc = Emotional Impact of Diabetes Treatment Questionnaire - Comparison; FSG = fasting serum glucose; GIEH = Global Impression of Emotional Health; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; IW-SP = Impact of Weight on Self Perception; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Clinical Trials Version; LDL = low-density lipoprotein; No-Hypo = No Hypoglycemia; SMBG = self-monitoring of blood glucose; T2D = type 2 diabetes; TG = triglyceride; VLDL = very-low-density lipoprotein.

Primary and key secondary endpoints are controlled for multiplicity.

Primary estimand

There will be 2 primary estimands evaluated in this study. Both of these estimands are described below.

- The treatment-regimen estimand is consistent with the US Prescribing Information. This estimand reflects efficacy when participants with T2D are treated in clinical practice and takes into account both tolerability and efficacy.
- The efficacy estimand focuses on the treatment effect if participants who underwent randomization continued to receive the study treatment without rescue for severe, persistent hyperglycemia, and/or prohibited medication. This estimand will be used in publications to inform prescribers and physicians.

The treatment-regimen estimand answers the following question of interest for the primary objective: What is the treatment difference in change from baseline in hemoglobin A1c (HbA1c) after 40 weeks of treatment in participants with T2D regardless of treatment discontinuation for any reason and regardless of initiation of antihyperglycemic rescue or prohibited treatment (except for tirzepatide or dulaglutide) or change in the background oral antihyperglycemic medication (OAM)?

This estimand is described by the following attributes:

- **Population** - Participants with T2D inadequately controlled with dulaglutide with or without a background of OAM. Further details can be found in Section 5 of the protocol.
- **Endpoint** - Change in HbA1c from baseline to Week 40 after randomization
- **Treatment condition** - The randomly assigned treatment regardless of treatment discontinuation with or without antihyperglycemic rescue or prohibited medication, except for tirzepatide or dulaglutide, or change in background OAM. Further details on study treatment and concomitant, including rescue, treatments can be found in Section 6 of the protocol.
- **Intercurrent events** - Intercurrent events of interest: “treatment discontinuation for any reason” and “initiation of antihyperglycemic rescue or prohibited treatment (except for tirzepatide or dulaglutide) or change in background OAM” are addressed by the treatment condition. Initiation of nonstudy tirzepatide or dulaglutide is addressed by the hypothetical strategy; that is, as if participants had not taken nonstudy tirzepatide or dulaglutide.
- **Population-level summary** - Difference in mean changes between treatment conditions

The efficacy estimand answers the following question of interest for the primary objective: What is the treatment difference in HbA1c change from baseline to Week 40 after randomization in participants with T2D regardless of changes in background OAM assuming that participants had stayed on treatment and not taken antihyperglycemic rescue medication?

This estimand is referred to as the efficacy estimand and is described by the following attributes:

- **Population** - Participants with T2D inadequately controlled with dulaglutide with or without background OAM. Further details can be found in Section 5 of the protocol.
- **Endpoint** - HbA1c change from baseline to Week 40 after randomization
- **Treatment condition** - The randomly assigned treatment regardless of changes to the background OAM. Further details on study treatment can be found in Section 6 of the protocol
- **Intercurrent event** - Intercurrent events of interest: "Treatment discontinuation for any reason" and "Initiation of antihyperglycemic rescue treatment" will be addressed using the following strategies:
  - had participants stayed on treatment (hypothetical strategy), and
  - had participants not taken antihyperglycemic rescue or prohibited medication, except for tirzepatide or dulaglutide (hypothetical strategy).
- **Population-level summary** - Difference in mean changes between treatment conditions.

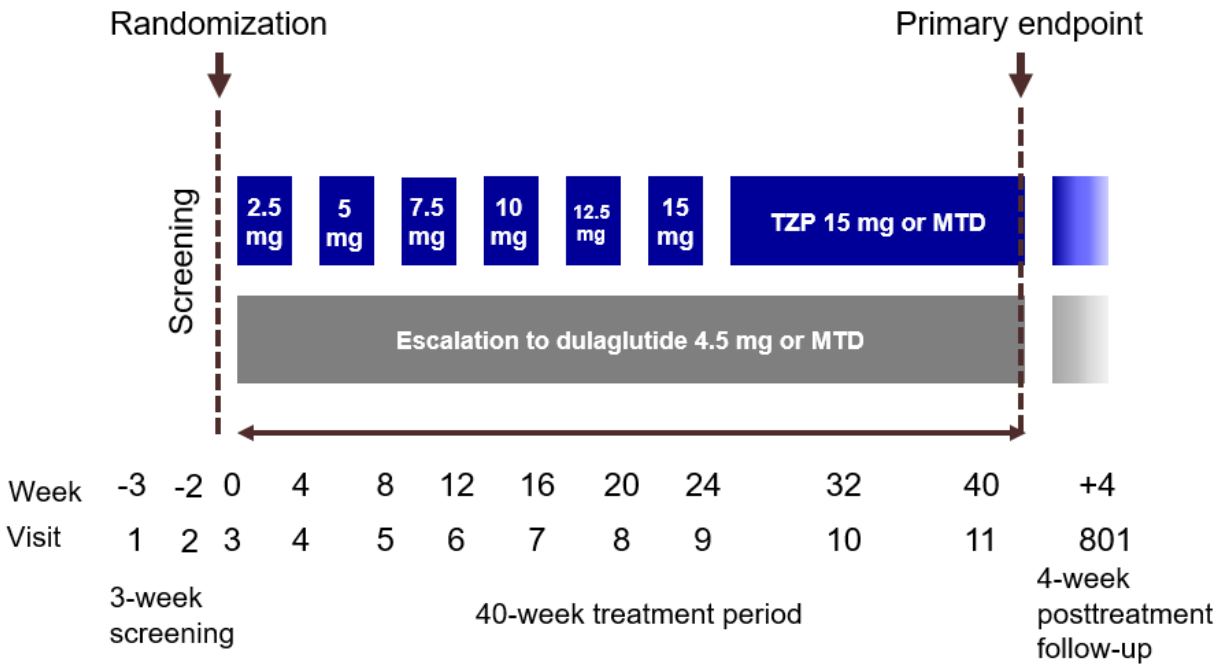
Both the efficacy and treatment-regimen estimands will be evaluated for the primary and key secondary objectives. The population, treatment condition, intercurrent events, and population-level summary specified above for each estimand for the primary objective will also apply to the key secondary objective.

1.2. Study Design

SURPASS-SWITCH is a Phase 4, randomized, open-label, active-controlled, parallel-group, multicenter, multinational trial to assess the efficacy and safety of tirzepatide 15 mg or MTD compared to dulaglutide 4.5 mg or MTD in participants with T2D.

This study will enroll participants with inadequately controlled T2D who are taking dulaglutide QW (0.75 mg or 1.5 mg) for at least 6 months prior to Visit 1, with or without (up to 3) background OAMs.

This study includes an approximate 3-week screening period, a 40-week treatment period, and a 4-week safety follow-up (Figure 1).



Abbreviations: MTD = maximum tolerated dose; TZP = tirzepatide.

Figure 1. Schema for Study GPIH.

Visit 1

Interested individuals will sign the informed consent form prior to initiating any procedure.

The investigator will review medical history and other inclusion and exclusion criteria prior to any diagnostic procedures.

Visit 2

The investigator will review the results of the screening laboratory measures to further assess participant eligibility.

For participants meeting all other eligibility requirements, a dilated fundoscopic examination will be scheduled and completed between Visit 2 and Visit 3.

Participants or caregivers, if applicable, will receive a blood glucose (BG) meter and clinical outcome assessment device.

Participants will be asked to monitor 7-point self-monitoring of BG profiles (prior to and 2 hours after the 3 main meals and at bedtime) on 2 nonconsecutive days during the 2 weeks preceding prespecified visits.

Participants should continue their diet and exercise routine and must not use any glucose-lowering treatment other than the OAM taken prior to the screening visit and should not change their dose or formulation.

Period II: treatment period

Visit 3 (Randomization)

All screening laboratory test results and screening dilated fundoscopic examination results must be reviewed for exclusion criteria prior to randomization and dosing at Visit 3.

This is the general flow for Visit 3:

- Participants arrive to the clinic in the fasting state.
- Study personnel confirm enrollment criteria.
- Participants are randomly assigned to an intervention group.
- Participants complete patient-reported outcome questionnaires before any other study procedures if the participant is not adversely affected by the fasting condition or complete after the participant has sufficiently recovered.
- Study personnel complete all required visit procedures, including the collection of vital signs, all baseline procedures, sample collection, and dispense study intervention.
- Participants self-administer study intervention at home. Administration should occur on the day (±3 days) of when the next dulaglutide administration is planned

Visit 4 through Visit 10

Participants complete all visit procedures described in the schedule of activities.

Visit 11 or early discontinuation visit

Participants unable or unwilling to continue the study for any reason will perform an early discontinuation (ED) visit. If the participant is discontinuing during an unscheduled visit or a scheduled visit, that visit should be performed as the ED visit.

Period III: safety follow-up visit

Visit 801

All participants who complete the treatment period or discontinue the study early (ED) are required to complete Visit 801, a safety follow-up visit, approximately 4 weeks after their last treatment or ED visit.

Method of treatment assignment

Participants who meet all entry criteria for enrollment will be centrally assigned to 1 of the study treatments arms at Visit 3. Assignment to treatment arms will be determined by a computer-generated random sequence using an interactive web-response system (IWRS).

Participants will be randomly assigned in a 1:1 ratio to tirzepatide or dose escalation of dulaglutide.

Participants will be stratified based on

- starting dulaglutide dose at screening (0.75 mg, 1.5 mg)
- region (North America, European Union)
- number of OAMs (0 to 1 or 2 to 3), and
- HbA1c (≤8.5% or >8.5%).

2. Statistical Hypotheses

The null hypothesis corresponding to the primary objective of this study is as follows:

- **$H_{1,0}$** - Switching from dulaglutide to 15 mg or MTD of tirzepatide is not superior to continuing and escalating to dulaglutide 4.5 mg or MTD with respect to HbA1c mean change from baseline at Week 40 in participants who entered the study on a nonmaximal dose of dulaglutide.

The null hypothesis corresponding to the key secondary objective is as follows:

- **$H_{2,0}$** – Switching from dulaglutide to 15 mg or MTD of tirzepatide is not superior to continuing and escalating to dulaglutide 4.5 mg or MTD with respect to weight mean change from baseline at Week 40 in participants who entered the study on a nonmaximal dose of dulaglutide.

Operationally the hypotheses will be evaluated by 2-sided tests.

2.1. Multiplicity Adjustment

The primary and the key secondary objectives will be evaluated using both the treatment-regimen and the efficacy estimands. Since the purpose for each estimand is different, multiplicity adjustment will be conducted for each estimand separately.

The statistical comparisons for the primary efficacy endpoint and the key secondary endpoint will be carried out in the hierarchical order. This means that statistically significant results for the comparison in the higher rank (primary) is required to initiate the testing of the next comparison in the lower rank (key secondary). Since a fixed sequence testing is used, each comparison will be tested at a significance level of 1-sided 0.025, and an overall alpha level of 1-sided 0.025 will be preserved (Westfall and Krishen, 2001).

3. Analysis Sets

This table defines the analysis populations and datasets for the purposes of analysis based on the estimands defined in Section 1.1.

Population/Analysis Set	Description
Screened population	All participants who signed informed consent
Randomized population	All participants who are randomly assigned to a treatment arm
Modified intent-to-treat population (mITT)	All randomly assigned participants who are exposed to at least 1 dose of study intervention. Participants will be analyzed according to the treatment they were randomly assigned to regardless of the treatment actually received.
Efficacy Analysis Set (EAS): This analysis set will be used to estimate the efficacy estimand for the primary and key secondary objectives	Data obtained during the treatment period from the mITT population excluding patients who were inadvertently enrolled, excluding data after permanent discontinuation of treatment or initiation of antihyperglycemic rescue medication or prohibited medication
Full Analysis Set (FAS): This analysis set will be used to estimate the treatment-regimen estimand for the primary and key secondary objectives	Data obtained during the treatment period from the mITT population excluding patients who were inadvertently enrolled, regardless of adherence to treatment or initiation of antihyperglycemic rescue medication or prohibited medication (excluding nonstudy tirzepatide or dulaglutide). Data after taking nonstudy tirzepatide or dulaglutide will be excluded.
Safety Analysis Set (SS): This analysis will be used to assess the safety of study treatment	Data obtained during the treatment period plus safety follow-up from the mITT population, regardless of adherence to treatment or initiation of rescue antihyperglycemic medication or prohibited medication.

4. Statistical Analyses

4.1. General Considerations

Statistical analysis of this study will be the responsibility of Lilly or its designee. The statistical analyses will be performed using SAS® Version 9.4 or higher.

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 change to the data analysis methods described in the protocol, and the justification for making the change, will be described in the SAP or clinical study report (CSR). Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (for example, too few events to justify conducting an analysis). Listings of events will be provided in such situations. Additional exploratory analyses of the data will be conducted as deemed appropriate without further changes made to the protocol or SAP, even after final database lock.

All tests of treatment effects will be conducted at a 2-sided alpha level of 0.05/1-sided alpha level of 0.025, unless otherwise stated, and all confidence intervals (CIs) will be given at a 2-sided (95%) level. In statistical summaries and analyses, all data will be analyzed by randomly assigned treatment assignment. Participants will be analyzed according to the treatment they were randomly assigned to, regardless of the treatment actually received.

Baseline is defined as the last nonmissing measurement recorded on or before the randomization visit, prior to first dose of treatment, unless otherwise specified. For laboratory values, baseline needs to be prior the first dose time.

Efficacy analyses will use the Efficacy Analysis Set (EAS) to evaluate the efficacy estimand and the Full Analysis Set (FAS) to evaluate the treatment-regimen estimand. Safety will be assessed using Safety Analysis Set (SS). Selected safety analyses may be conducted after excluding data on rescue therapy or data after starting another antihyperglycemic medication.

End of study participation for a patient will be the earliest of date of death, date of withdrawal from further participation in the study, or date of safety follow-up visit (Visit 801). For patients considered to be lost to follow-up, end of study participation will be the date of lost to follow-up reported by the investigator. Patient data included in the database after the last date of study participation (earliest of date of death, date of ED or date of safety follow-up) will be excluded from statistical analysis. Listings of such data may be provided.

Summary statistics for continuous measures may include sample size, mean, standard deviation, median, minimum, and maximum. The analysis model to make comparisons between treatment groups relative to continuous measurements assessed for specific time will be an analysis of covariance (ANCOVA) and for over time will be mixed-effects model for repeated measures (MMRM).

The Kaplan-Meier method will be used for estimation of cumulative event-free survival rates over time, and Cox proportional hazards regression analysis will be used to compare hazard rates among treatments.

Summary statistics for categorical measures, including categorized continuous measures, will include sample size, frequency, and percentages. Logistic regression will be used to examine the treatment difference in binary efficacy outcomes when 1 or more covariates are included in the model. Otherwise, Fisher's exact text will be used. The negative binomial regression model will be used for the treatment comparison of discrete count measures if deemed appropriate.

4.2. Participant Dispositions

Reasons for screen failures as reported by investigators will be summarized.

A listing of final study disposition and a listing of randomized treatment assignment (planned treatment) for all randomized patients will be provided. Final study disposition and study treatment disposition for all randomized patients will be summarized by planned study treatment.

The number and percentage of participants prematurely discontinuing study treatment and study prior to the 40-week visit will be provided by study treatment. Reasons for prematurely discontinuing study treatment and study prior to the 40-week visit will be provided by study treatment.

4.3. Primary Endpoint Analysis

The primary endpoint for this study is change from baseline in HbA1c to the 40-week visit (Visit 11). This endpoint will be used to evaluate the primary objective of the study for both the treatment regimen and the efficacy estimands (Section 1.1).

The primary efficacy measure will be mean change in HbA1c (percentage and mmol/mol) from baseline (postbaseline to baseline). Both HbA1c values as well as change from baseline in HbA1c will be summarized by treatment and nominal visit (week). If scheduled HbA1c data at the primary endpoint visit are not available, unscheduled HbA1c data collected for the primary endpoint visit will be included in the analysis.

4.3.1. Primary Analysis Relative to the Treatment-Regimen Estimand

The primary analysis relative to the treatment-regimen estimand will be conducted utilizing HbA1c data in the FAS at baseline and at the 40-week visit with the aid of an ANCOVA model. ANCOVA model for change from baseline in HbA1c at Week 40 with terms includes treatment (tirzepatide and dulaglutide), number of background OAMs (0 to 1 or 2 to 3), dulaglutide dose at screening (0.75 mg or 1.5 mg), region (North America, European Union) as fixed effect, and baseline HbA1c as a covariate. The ANCOVA analysis will be conducted with multiple imputation of missing primary measure and statistical inference over multiple imputation of missing data guided by Rubin (1987).

For the purpose of the treatment-regimen estimand, missing HbA1c values at the 40-week visit will be imputed based on observed data in the same treatment arm from participants who had their efficacy assessed after ED of study intervention (retrieved dropouts). In cases where there are not enough retrieved dropouts to provide a reliable imputation model (in other words, the model implemented by the SAS program does not converge), an alternative multiple imputation method with reference to the baseline values (return to baseline multiple imputation) will be used

as the primary analysis relative to the treatment-regimen estimand. If value of the imputed HbA1c change from baseline is <-6.0% or >6.0%, that value will be set to -6.0% or 6.0%, respectively, to avoid unrealistic imputed values.

With the aid of the ANCOVA analysis, p-values, and 2-sided 95% CIs for mean change in HbA1c from baseline to the 40-week visit will be derived and summarized for the tirzepatide 15 mg or MTD arm compared to dulaglutide 4.5 mg or MTD arm.

4.3.2. Primary Analysis Relative to the Efficacy Estimand

The primary analysis relative to the "efficacy" estimand will be conducted using HbA1c data in the EAS from baseline through the 40-week visit with the aid of MMRM. Restricted maximum likelihood (REML) will be used to obtain model parameter estimates, and the Kenward-Roger option will be used to estimate denominator degrees of freedom. The response variable of the MMRM model will be change in HbA1c values from baseline obtained at each scheduled postbaseline visit. The independent variables of the MMRM model are treatment, visit, treatment-by-visit interaction, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening, and region as fixed effects, and baseline HbA1c as covariate. Missing data will be addressed by the MMRM model. No explicit imputation methods for missing data will be employed.

An unstructured covariance matrix will be used to model the within-patient errors. If this model fails to converge, the following covariance structures will be tested in the following order:

1. Heterogeneous Toeplitz
2. Heterogeneous First Order Autoregressive
3. Heterogeneous Compound Symmetry
4. Toeplitz
5. First Order Autoregressive
6. Compound Symmetry

The first covariance structure that converges will be used. The resulting least squares mean estimate of mean change from baseline in HbA1c will be summarized by visit and by study treatment.

With the aid of the MMRM analysis, p-values and 2-sided 95% CIs for mean change in HbA1c from baseline to the 40-week visit will be derived and summarized for the tirzepatide 15 mg or MTD arm compared to dulaglutide 4.5 mg or MTD arm

4.4. Secondary Endpoints Analysis

4.4.1. Key Secondary Endpoint

The endpoint corresponding to secondary study objective subject to type 1 error rate control is specified in Section 1.1 under "Key Secondary (Controlled for Type 1 error) endpoints".

The null hypotheses corresponding to the key secondary objective can be found in Section 2.

The key secondary objective will be evaluated based on the treatment-regiment and the efficacy estimands (Section 1.1), similar to the primary objective.

4.4.1.1. Mean Changes in Body Weight (kg and lb) from Baseline at 40-week Visit

The analysis for change in body weight (kg and lb) from baseline (postbaseline to baseline) at Week 40 will be conducted in a manner similar to the primary analysis in Section 4.3.1 (treatment-regimen estimand) and Section 4.3.2 (efficacy estimand) with the following differences:

- For the treatment regimen estimand, the ANCOVA model (Section 4.3.1), baseline category for HbA1c (≤8.5%, >8.5%) will be used as a fixed factor in place of baseline HbA1c as a covariate, and baseline weight will be used as an additional covariate in the statistical model. Imputation of missing values at Week 40 will be done in a similar manner as described in Section 4.3.1. If value of the imputed weight change from baseline is <-50 kg or >50 kg, that value will be set to -50 kg or 50 kg, respectively, to avoid unrealistic imputed values. Similar imputed values (equivalent to kg) will be used for body weight change (lb) from baseline endpoints, to avoid unrealistic imputed values.
- For the efficacy estimand, the MMRM model (Section 4.3.2) will be updated by adding the baseline category for HbA1c (≤8.5%, >8.5%) as a fixed factor in place of baseline HbA1c and baseline weight as an additional covariate. Missing data will be addressed by the MMRM model, and no explicit imputation methods will be used.

4.4.2. Additional Secondary Endpoints

Additional secondary endpoints except for endpoints specified below will use the efficacy estimand using EAS and will be summarized by treatment and nominal visit.

- change in fasting serum glucose (FSG) (via central lab) from baseline
- HbA1c <5.7%, ≤6.5%, and <7%
- weight loss from baseline of ≥5%, ≥10%, and ≥15%
- change in Waist Circumference from baseline
- change in body mass index (BMI) from baseline
- a composite endpoint “HbA1c ≤ 6.5% & Weight Loss ≥10% & No-Hypo” defined as
  - HbA1c ≤ 6.5%
  - weight loss ≥10%, and
  - no hypoglycemia, defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia

The endpoints specified above will be analyzed using both treatment regimen and efficacy estimands similar to primary analysis specified in Section 4.3. Unless otherwise specified,

additional secondary endpoints are not subject to type 1 error rate control. Some parameters may be log transformed, if necessary.

4.4.2.1. Analysis for Continuous Outcomes

The analysis to make comparisons between treatment groups for additional secondary continuous outcomes relative to the "efficacy" estimand will be conducted similar to the primary efficacy analysis in Section 4.3.2. The MMRM model will be used for endpoints that have repeated measures over time and the ANCOVA model will be implemented to assess for specific time points. The MMRM model will be updated by adding the baseline category for HbA1c (≤8.5%, >8.5%) as a fixed factor in place of baseline HbA1c, and corresponding baseline values as an additional covariate. Missing data will be addressed by the MMRM model, and no explicit imputation methods will be used. The ANCOVA model (Section 4.3.1) will be updated by adding the baseline category for HbA1c (≤8.5%, >8.5%) as a fixed factor in place of baseline HbA1c, and corresponding baseline values (if available) will be used as an additional covariate in the statistical model.

For the analysis of FSG, waist circumference, and BMI using the treatment regimen estimand, the ANCOVA model (Section 4.3.1) will be implemented with baseline category for HbA1c (≤8.5%,>8.5%) as a fixed factor in place of baseline HbA1c as a covariate and baseline values will be used as an additional covariate in the statistical model.

4.4.2.2. Analysis for Binary Outcomes

The analysis for binary outcomes relative to the "efficacy" estimand will be performed using EAS with missing values imputed by the predicted value from the MMRM model specified in Section 4.3.2, and dichotomized data will then be derived based on continuous imputed values. After dichotomizing, the data is analyzed using a logistic regression model.

In addition, analysis will be conducted utilizing data using EAS from baseline through the 40-week visit with the aid of a longitudinal logistic regression with repeated measurements with treatment, visit, treatment-by-visit interaction, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening, region, baseline category for HbA1c (≤8.5%,>8.5%) as fixed effect, and baseline of the corresponding variable as covariate. In case the longitudinal logistic model does not converge due to small number of events, logistic regression will be utilized at nominal visits.

- For analysis of HbA1c binary endpoints, the logistic model will include treatment, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening, region as fixed effects, and baseline HbA1c as covariates.
- For analysis of weight related binary endpoints, the logistic model will include treatment, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening, region and baseline category for HbA1c (≤8.5%,>8.5%) as fixed effects, and baseline weight as covariates.

- For analysis of composite binary endpoints, the logistic regression model will include treatment, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening and region as fixed effects, and baseline HbA1c, and baseline weight as covariates

The analysis for binary outcomes relative to the "treatment-regimen" estimand will be performed using FAS at baseline and at the 40-week visits with the aid of a logistic regression with multiple imputation of missing data at the 40-week visit (see Section 4.4.1 for details), and dichotomized data will then be derived based on continuous imputed values. After dichotomizing, the data is analyzed using a logistic regression model. Statistical inference over multiple imputations will be guided by Rubin (1987).

- For analysis of HbA1c binary endpoints, the logistic model will include treatment, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening, region as fixed effects, and baseline HbA1c as covariates.
- For the analysis of weight-related binary endpoints, the logistic model will include treatment, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening, region, baseline category for HbA1c (≤8.5%,>8.5%) as fixed effects, and baseline weight as covariates.
- For analysis of composite binary endpoints, the logistic regression model will include treatment, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening and region as fixed effects, and baseline HbA1c and baseline weight as covariates.

Table 2. Additional Secondary Efficacy Analyses Not Controlled for Type I Error

Objective	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to Section	Additional Information
Additional Secondary Analyses			
To assess the clinical effect of tirzepatide compared to dulaglutide at Week 40	Proportion of patients achieving an HbA1c target value of <5.7%, from baseline	Logistic model for both treatment regimen and efficacy estimands specified in Section 4.4.2.2	None
	Proportion of patients achieving an HbA1c target value of ≤6.5%, from baseline	Logistic model for both treatment regimen and efficacy estimands specified in Section 4.4.2.2	None
	Proportion of patients achieving an HbA1c target value of <7%, from baseline	Logistic model for both treatment regimen and efficacy estimands specified in Section 4.4.2.2	None
	Proportion of patients who achieved weight loss of ≥5%, from baseline	Logistic model for both treatment regimen and efficacy estimands specified in Section 4.4.2.2	None

Objective	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to Section	Additional Information
	Proportion of patients who achieved weight loss of ≥10%, from baseline	Logistic model for both treatment regimen and efficacy estimands specified in Section 4.4.2.2	None
	Proportion of patients who achieved weight loss of ≥15%, from baseline	Logistic model for both treatment regimen and efficacy estimand specified in Section 4.4.2.2	None
	Proportion of patients who achieved composite endpoint ("HbA1c ≤6.5% & Weight Loss ≥10% & No-Hypo"), defined as • HbA1c ≤6.5% • weight loss ≥10%, and • no hypoglycemia, defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia	Logistic model for both treatment regimen and efficacy estimand specified in Section 4.4.2.2	None
	Change from baseline in fasting serum glucose	ANCOVA model in Section 4.4.2.1 for treatment regimen estimand and MMRM model in Section 4.4.2.1 for efficacy estimand	Use baseline fasting serum glucose as a covariate. LSM estimates through 40 weeks will be plotted by study treatment.
	Change from baseline in waist circumference	ANCOVA model in Section 4.4.2.1 for treatment regimen estimand and MMRM model in Section 4.4.2.1 for efficacy estimand	Use baseline waist circumference as a covariate.
	Change from baseline in BMI	ANCOVA model in Section 4.4.2.1 for treatment regimen estimand and MMRM model in Section 4.4.2.1 for efficacy estimand	Use baseline BMI as a covariate.
	Change in IWQOL-Lite-CT Physical Functioning domain score from baseline	ANCOVA model in Section 4.4.2.1	Use baseline IWQOL-Lite-CT Physical Functioning domain score as a covariate.

Abbreviations: ANCOVA = analysis of covariance; BG = blood glucose; BMI = body mass index; HbA1c = hemoglobin A1c; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Clinical Trials Version; LSM = least squares mean; MMRM = mixed-effect model of repeated measures; No-Hypo = No Hypoglycemia.

4.5. Exploratory Endpoints Analysis

Other exploratory (prespecified/nonprespecified in the protocol) endpoints except for endpoints specified below will use the efficacy estimand and will be summarized by treatment and nominal visit. Unless otherwise specified, missing data will not be explicitly imputed, and assessments are not subject to the type 1 error rate control. Some parameters may be log transformed, if necessary.

The endpoint specified below will be analyzed using both treatment regimen and efficacy estimand similar to primary analysis specified in Section 4.3:

- percent change from baseline in body weight.

Table 3. Exploratory Efficacy Analyses Not Controlled for Type I Error

Objective	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to Section	Additional Information
Exploratory Objectives			
To assess the clinical effect of tirzepatide compared to dulaglutide at Week 40	Change from baseline in lipid parameters (Total-Cholesterol, HDL-C, LDL-C, VLDL-C, TG)	MMRM model in Section 4.4.2.1	Use corresponding baseline lipid parameters as a covariate.
	Change in daily average 7-point self-monitored blood glucose profiles	ANCOVA model in Section 4.4.2.1	Use baseline SMBG parameter as a covariate.
	Change in IWQOL-Lite-CT (Total, Physical, and Psychosocial) from baseline	ANCOVA model in Section 4.4.2.1	Use baseline IWQOL-Lite-CT (Total, Physical, and Psychosocial) as a covariate.
	Change in Impact of Weight on Self Perception (IW-SP)	ANCOVA model in Section 4.4.2.1	Use baseline IW-SP score as a covariate.
	Change in ability to Perform Physical Activities of Daily Living (APPADL) from baseline	ANCOVA model in Section 4.4.2.1	Use baseline APPADL score as a covariate.
	EIDTQc at Week 40	ANCOVA model in Section 4.4.2.1	None
	Change in GIEH-1 week recall value from baseline	ANCOVA model in Section 4.4.2.1	Use baseline GIEH-1 week recall value as a covariate.
	GIEH-1 since study start recall value at Week 40	ANCOVA model in Section 4.4.2.1	None

Abbreviations: ANCOVA = analysis of covariance; EIDTQc = Emotional Impact of Diabetes Treatment Questionnaire - Comparison; GIEH = Global Impression of Emotional Health; HDL-C = high-density lipoprotein cholesterol; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Clinical Trials Version; LDL-C = low-density lipoprotein cholesterol; MMRM = mixed-effects model for repeated measures; SMBG = self-monitoring of blood glucose; TG = triglyceride; VLDL-C = very-low-density lipoprotein cholesterol.

Note: Exploratory efficacy analyses are not controlled for type I error.

4.5.1. Exploratory Analyses Not Specified in the Protocol

Table 4. Nonprotocol-Specified Exploratory Efficacy Analyses Not Controlled for Type I Error

Objective	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to Section	Additional Information
To assess the clinical effect of tirzepatide compared to dulaglutide at Week 40	Percent change from baseline in body weight	ANCOVA model in Section 4.4.1.1 for treatment regimen estimand and MMRM model in Section 4.4.1.1 for efficacy estimand	None
	Proportion of patients who achieved composite endpoint ("HbA1c <5.7% & Weight Loss ≥15% & No-Hypo"), defined as • HbA1c <5.7% • weight loss ≥15%, and • no hypoglycemia, defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia	Logistic model specified in Section 4.4.2.2	None

Abbreviations: ANCOVA = analysis of covariance; BG = blood glucose; HbA1c = hemoglobin A1c; No-Hypo = No Hypoglycemia.

4.6. Safety Analyses

Safety assessments will be done using the SS (See Section 3) irrespective of adherence to study intervention or initiation of rescue antihyperglycemic medication or prohibited medication, unless indicated otherwise. AEs will be coded from the actual term using the Medical Dictionary for Regulatory Activities (MedDRA) and reported with Preferred Terms (PTs) and System Organ Class (SOC). Selected notable AEs of interest may be reported using High-Level Terms (HLTs) or Standardized MedDRA Queries (SMQs). Summary statistics will be provided for incidence of treatment-emergent AEs (TEAEs), serious AEs (SAEs), and study discontinuation due to AEs, study intervention discontinuation due to AEs, or deaths from the time of first dose through the end of safety follow-up. Counts and proportions of participants experiencing AEs will be reported for each treatment group, and Fisher's exact test will be used to compare the treatment groups.

Where necessary, for selected continuous safety parameters, the mean change from baseline at Week 40 will be assessed via MMRM using REML. Data from scheduled visits will be utilized for this analysis unless specified otherwise. The model for the analysis during the 40-week treatment period will include treatment, visit, treatment-by-visit interaction, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening, and region as fixed effects, and baseline value of the safety parameter as a covariate. To model the covariance structure within participants, the unstructured covariance matrix will be used. If this model fails to converge, the covariance structures specified in Section 4.3.2 will be tested in order until met convergence. If the data does not warrant the MMRM model, then ANCOVA model will be conducted. Where

necessary, the rate of events will be analyzed using a generalized linear MMRM assuming the number of events follow a negative binomial distribution and with treatment as a fixed effect.

4.6.1. Adverse Events

AEs will be coded from the actual term using MedDRA and reported with PTs and SOC.

A TEAE is defined as medical occurrence that emerges during a defined treatment period, having been absent pretreatment, or worsens relative to the pretreatment state, and does not necessarily have to have a causal relationship with this treatment. The maximum severity for each Low-Level Term (LLT) during the baseline 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.

The percentages of patients with TEAEs will be summarized by treatment using MedDRA PT nested within SOC, and Fisher's exact test will be used to compare the treatment groups at both the SOC and PT levels. Events will be ordered by decreasing frequency within SOC. For events that are sex-specific, the denominator and computation of the percentage will include only patients from the given sex.

Overview of the number and percentage of patient who experienced a TEAE, SAE, death, discontinued from study treatment or study due to an AE, and relationship to study drug, will be summarized by treatment.

The percentages of patients with TEAEs, overall and common (common TEAEs occurred in ≥5% of patients before rounding), will be summarized by treatment using MedDRA PT. Events will be ordered by decreasing frequency.

The percentages of patients with TEAEs by maximum severity will be summarized by treatment using MedDRA PT. For each patient and TEAE, the maximum severity for the MedDRA PT is the maximum postbaseline severity observed from all associated LLTs mapping to the MedDRA PT. The maximum severity will be determined based on the nonmissing severities. If all severities are missing for the defined postbaseline period of interest, it will show as missing in the table. Only counts and percentages will be included for the TEAEs by maximum severity.

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

- deaths
- SAEs
- pregnancy, and
- permanent discontinuations of study treatment due to AEs.

4.6.1.1. Deaths

A listing of all deaths will be provided. The listing will include patient identification (ID) including the treatment, site number, date of death, age at the time of enrollment, gender, MedDRA PT of associated AE, time from first dose of study drug to death, time from last dose of study drug to death (if patient had discontinued study drug), cause of death as reported by investigator, cause of death as adjudicated by Clinical Endpoint Committee.

4.6.1.2. Serious Adverse Events

The number and percentage of patients who experienced an SAE (including deaths and SAEs temporally associated or preceding deaths) will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC. A listing of all SAEs will be provided. Listing will include but not limited to treatment, patient ID including the site number, treatment group, date of event, age at the time of enrollment, gender, MedDRA SOC and PT, severity, action taken, outcome, relationship to study drug, time from first dose of study drug to the event, and event duration.

4.6.1.3. Discontinuation from Study Due to Adverse Event

The number and percentage of patients who prematurely discontinue the study due to an adverse event (AE) will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC. A listing of all the discontinuation from study due to AE will be provided.

4.6.1.4. Discontinuation from Study Treatment 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. A listing of all the discontinuation from study treatment due to AE will be provided.

4.6.2. Special Safety Topics

4.6.2.1. Hypoglycemic Events

Definitions of different categories of hypoglycemic events are mentioned below.

Level 1

Glucose <70 mg/dL (<3.9 mmol/L) and ≥54 mg/dL (≥3.0 mmol/L)

Level 1 hypoglycemia can alert a person to take action such as treatment with fast-acting carbohydrates. Providers should continue to counsel participants to treat hypoglycemia at this glucose alert value.

Level 2

Glucose <54 mg/dL (<3.0 mmol/L)

Level 2 hypoglycemia is also referred to as documented or BG-confirmed hypoglycemia with glucose <54 mg/dL (<3.0 mmol/L). This glucose threshold is clinically relevant regardless of the presence or absence of symptoms of hypoglycemia.

Level 3 Severe

A severe hypoglycemic event is characterized by altered mental or physical status requiring assistance of another person to actively administer carbohydrate, glucagon, or other resuscitative actions for the treatment of hypoglycemia.

The determination of an episode of severe hypoglycemia is made by the investigator based on the medical need of the participant to have required assistance and is not predicated on the report of a participant simply having received assistance.

Examples of severe hypoglycemia in adults are

- altered mental status and the inability to assist in their own care
- semiconscious or unconscious, or
- coma with or without seizures.

Severe hypoglycemia: A severe hypoglycemic event is characterized by altered mental or physical status requiring assistance of another person to actively administer carbohydrate, glucagon, or other resuscitative actions for the treatment of hypoglycemia.

Nocturnal hypoglycemia: Nocturnal hypoglycemia is a hypoglycemia event, including severe hypoglycemia, which **occurs at night** and presumably during sleep.

To avoid duplicate reporting, all consecutive BG values occurring within a 1-hour period may be considered to be a single hypoglycemic event and the one with the lowest BG value can be selected to be the representative. If two records with the same lowest values within a hour, the earlier occurrence will be selected.

Statistical summaries and analyses will exclude hypoglycemic events occurring after initiation of a new antihyperglycemic therapy. For severe hypoglycemia and level 2 hypoglycemia incidence as well as rate per patient year of exposure will be provided by treatment at specified time intervals. The incidence of hypoglycemic event will be analyzed using logistic regression with treatment, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening, baseline HbA1c category (≤8.5%,>8.5%), and region as fixed effects. The rate of hypoglycemic episodes per patient year may be analyzed using a generalized linear MMRM assuming the number of hypoglycemic episodes follows a negative binomial distribution with the mean modeled using region, number of background OAMs (0 to 1, 2 to 3), dulaglutide dose at screening, baseline HbA1c category (≤8.5%,>8.5%), and treatment as fixed effects if data warrants. The logarithm of years in specified time interval will be adjusted as an offset to account for possible unequal

treatment duration in the specified time interval between patients. When the number of hypoglycemic events is less than 10, the listing of hypoglycemic events will be provided instead.

4.6.2.2. Severe Persistent Hyperglycemia

A summary statistic of initiation of rescue therapy in response to severe, persistent hyperglycemia will be provided by treatment. A listing of participants who initiated rescue therapy will be provided.

4.6.2.3. Pancreatitis

If data warrants, summaries of adjudicated and investigator-reported pancreatic events will be provided by treatment. Determination of investigator-reported events will be through the predefined SMQ search for acute pancreatitis and MedDRA PT of pancreatitis chronic. Detailed searching criteria can be found in Section 6.5 (Appendix 5).

4.6.2.3.1. Pancreatic Hyperenzymemia

Observed pancreatic enzyme data (p-amylase and lipase) will be summarized and listed by treatment and nominal visit.

Additionally, the number and proportion of patients with maximum postbaseline pancreatic enzyme values exceeding the following thresholds will be provided by maximum baseline pancreatic enzyme value (≤1×upper limit of normal (ULN), >1×ULN), and treatment: ≤1×ULN, (>1 to ≤3) × ULN, (>3 to ≤5) × ULN, (>5 to ≤10) × ULN, >10 × ULN.

4.6.2.4. Thyroid C-Cell Hyperplasia and C-Cell Neoplasms

Treatment-emergent thyroid C-cell hyperplasia and C-cell neoplasms will be identified using predefined MedDRA HLTs of thyroid neoplasms malignant and PT of thyroid C-cell hyperplasia. Detailed searching criteria can be found in Section 6.5 (Appendix 5). A summary and a listing by treatment and PT will be provided.

4.6.2.4.1. Calcitonin

Observed calcitonin data will be summarized by treatment and nominal visit. Additionally, the number and proportion of patients with a maximum postbaseline calcitonin value exceeding the following thresholds will be provided by treatment and maximum baseline calcitonin value (≤20 ng/L, >20 ng/L to ≤35 ng/L, >35 ng/L): ≤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.

4.6.2.5. Major Adverse Cardiovascular Events

Deaths and nonfatal cardiovascular events AEs will be adjudicated by a committee of physicians external to Lilly with cardiology expertise. The nonfatal cardiovascular events AEs to be adjudicated include the following:

- myocardial infarction
- hospitalization for unstable angina
- hospitalization for heart failure

- coronary interventions, such as coronary artery bypass graft or percutaneous coronary intervention, and
- cerebrovascular events, including cerebrovascular accident (stroke) and transient ischemic attack.

The number and proportion of patients with major adverse cardiovascular events will be reported by treatment. Listing of deaths, myocardial infarctions, strokes, and hospitalization for unstable angina confirmed by an independent clinical endpoint committee will be provided. The dates of randomization, event, first dose and last dose of study intervention, and time from randomization to event will be listed.

4.6.2.6. Arrhythmias and Cardiac Conduction Disorders

The AE database will be searched using predefined SMQs or MedDRA HLTs to identify events consistent with supraventricular arrhythmias and cardiac conduction disorders. Detailed searching criteria can be found in Section 6.5 (Appendix 5). Incidence of the resulting TEAEs will be summarized by treatment and PT within SMQ and HLT.

4.6.2.7. Hypersensitivity Events

Hypersensitivity reactions and related information will be summarized by treatment. Two main analyses are performed:

- **Potential Immediate Hypersensitivity –** Analysis of TEAEs occurring from the start of study drug administration up to 24 hours after the end of study drug administration. For events without the hypersensitivity electronic case report form (eCRF), only date (no time) information are collected, the events occurred on the same date as the study drug injection date will be included.
- **Potential Nonimmediate Hypersensitivity –** Analysis of TEAEs occurring more than 24 hours after the end of study drug administration, but prior to subsequent study drug administration.

Summaries of all potential hypersensitivity reactions will be generated by PT with decreasing frequency by treatment. Detailed searching criteria can be found in Section 6.5 (Appendix 5).

4.6.2.8. Injection Site Reactions

Injection site reactions, incidence, and related information reported via the "Injection Site Reactions" eCRF will be summarized by treatment. Information to be summarized includes the timing of the reaction relative to study drug administration, and characteristics of the injection site reaction: erythema, induration, pain, pruritis, and edema. Patient based and event-based summaries will be created.

Additionally, potential injection site reactions will be searched by predefined MedDRA HLTs of injection site reactions, administration site reactions, and infusion related reactions. Detailed searching criteria for injection site reaction events can be found in Section 6.5 (Appendix 5). The PT will be used for summary by treatment within each HLT category.

4.6.2.9. Diabetic Retinopathy Complications

If data warranted, a listing of unscheduled visits only will be provided. Any TEAE suspected of worsening retinopathy triggers a follow-up dilated fundoscopic exam. A summary of TEAEs suspected of worsening retinopathy and a summary of the results of the follow-up dilated fundoscopic exam will be summarized by treatment and PT, if numbers allow it. Otherwise, listings would be provided.

The cases with repeated fundoscopy during the course of the trial, based on clinical suspicion of worsening retinopathy that have either findings of de novo retinopathy or progression of retinopathy, and severe/SAEs from the PTs defined in searching criteria in Section 6.5 (Appendix 5) will be summarized, if numbers allow it. Otherwise, listings will be provided.

4.6.2.10. Hepatic Safety

4.6.2.10.1. Hepatobiliary Disorders

The AE database will be searched using SMQs to identify events consistent with hepatobiliary disorders. Detailed searching criteria can be found in Section 6.5 (Appendix 5). A summary by treatment and PT within SMQ will be provided.

4.6.2.10.2. Acute Gallbladder Disease

The AE database will be searched using predefined SMQs to identify events consistent with acute gallbladder diseases. Detailed searching criteria for these AEs can be found in Section 6.5 (Appendix 5). A summary by treatment and PT within SMQ will be provided.

4.6.2.10.3. Liver Enzymes

Analyses for laboratory analyte measurements are described in Section 4.6.4 in the protocol. This section describes additional analyses of liver enzymes. In addition, the following will be provided by treatment group:

- A shift table of maximum to maximum alanine aminotransferase (ALT) measurement from baseline (≤1 × ULN, > 1 × ULN) to postbaseline with the following categories - ≤1 × ULN, >1 to <3 × ULN, ≥3 to <5 × ULN, ≥5 to <10 × ULN, ≥10 × ULN
- A shift table of maximum to maximum aspartate transaminase (AST) measurement from baseline (≤1 × ULN, > 1 × ULN) to postbaseline with the following categories - ≤1 × ULN, >1 to <3 × ULN, ≥3 to <5 × ULN, ≥5 to <10 × ULN, ≥10 × ULN.
- Shift tables of maximum to maximum total bilirubin and direct bilirubin from baseline to postbaseline with the following categories - ≤1 × ULN, > 1 to <2 × ULN, ≥2 × ULN.
- Shift tables of serum alkaline phosphatase from baseline to postbaseline with the following categories - ≤1 × ULN, >1 to <2 × ULN, ≥2 × ULN.

Maximum baseline will be the maximum nonmissing observation in the baseline period. The maximum postbaseline value will be the maximum nonmissing value from the postbaseline period. Planned and unplanned measurements will be included.

4.6.2.11. Gastrointestinal Safety

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

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

The PTs in the gastrointestinal SOC will be used to identify gastrointestinal AEs. The incidence of the resulting TEAEs will be summarized by treatment and PT.

4.6.3. Vital Signs

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 using REML model will be used to analyze the changes from baseline in vital signs at all scheduled postbaseline visits. The model will include treatment, visit, and treatment-by-visit interaction, number of background OAMs (0 to 1, 2 to or le3), dulaglutide dose at screening region, baseline HbA1c (≤8.5%, >8.5%), region as fixed effects, and baseline value of the dependent variable as a covariate.

Counts and percentages of patients with treatment-emergent abnormal sitting systolic blood pressure (BP), sitting diastolic BP, and pulse will be presented by treatment. The criteria for identifying patients with treatment-emergent vital sign abnormalities are stated in Table 5.

Table 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	≥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.

4.6.4. Clinical Laboratory Evaluation

All laboratory data will be reported in the International System of Units and Conventional Units. Values that are outside of reference ranges will be flagged as high (H) or low (L) in the listings. Descriptive summaries by treatment and by nominal visit will be provided for the baseline and postbaseline values as well as the change from baseline values for selected measurements.

Observed and change from baseline values for selected measurements for each visit will be displayed graphically for patients who have both a baseline and a postbaseline planned measurement. Unplanned measurements will be excluded from graphs.

Shift tables will be produced for selected measurements. A shift table will include unplanned measurements. The shift table will include the number and percentage of patients within each baseline category (low, normal, high, or missing) versus each postbaseline category (low, normal, high, or missing) by treatment. The proportion of patients shifted will be compared between treatments.

A listing of abnormal findings will be created for laboratory analyte measurements. The listing will include patient ID, treatment group, laboratory collection date, study day, analyte name, and analyte finding.

4.6.5. Device Product Complaints

A product complaint is any written, electronic, or oral communication that alleges deficiencies related to the identity, quality, durability, reliability, safety, effectiveness, or performance of a study intervention. When the ability to use the study intervention safely is impacted.

Product complaints related to study interventions used in clinical studies are collected in order to ensure the safety of participants, monitor quality, and to facilitate process and product improvements.

Investigators will instruct participants to contact the site as soon as possible if he or she has a product complaint or problem with the study intervention so that the situation can be assessed.

An event may meet the definition of both a product complaint and an AE/SAE. In such cases, it should be reported as both a product complaint and as an AE/SAE.

A listing of all device product complaints, inclusive of device product complaints that lead to an AE or that could have led to an SAE had intervention not been taken will be provided. Additional summaries will be provided as deemed appropriate.

4.7. Other Analyses

4.7.1. Subgroup Analyses

Subgroup analyses of the primary endpoint (change from baseline in HbA1c) and key secondary endpoint (change from baseline in weight) will be made to assess consistency of the intervention effect across the following subgroups:

- Age group - <65 years, ≥65 years
- Sex – female, male
- Race - white, black, other
- Ethnicity - Hispanic, non-Hispanic
- Duration of diabetes - ≤5 years, >5 to ≤10 years, >10 years

- Baseline HbA1c - ≤8.5%, >8.5%
- Region - North America, European Union
- Baseline dulaglutide doses - 0.75 mg and 1.5 mg
- Baseline duration of dulaglutide doses - <1 and ≥1 year
- Baseline BMI category - <27, ≥27, <30, ≥30 and <35, ≥35 kg/m^2

4.8. Interim Analyses

No interim analyses are planned for this study. If an unplanned interim analysis is deemed necessary for reasons other than a safety concern, the protocol must be amended.

4.9. Changes to Protocol-Planned Analyses

The only change to the protocol planned analyses that is proposed in the SAP is related to the description of the analyses data sets in Section 3. In this section, it is clarified that data after taking nonstudy tirzepatide or dulaglutide will be excluded from the FAS.

5. Sample Size Determination

Approximately 250 participants will be randomly assigned in a 1:1 ratio to each study arm. This sample size provides at least 90% power to detect a difference between treatment arms at Week 40 irrespective of adherence to treatment, changes in background OAMs or introduction of rescue antihyperglycemic medication. This sample size is based on an assumed treatment difference of 0.58% with a common standard deviation of 1.3%, using a 2-group t-test and a 0.05 two-sided significance level. The treatment difference is based on the following assumptions for the dose distribution for maximum dose or MTD:

- Tirzepatide MTD will be comprised of the following doses:
    - 15 mg dose - 80% of the participants
    - 10 mg dose - 15% of the participants
    - 5 mg dose - 5% of the participants
- Dulaglutide MTD will be comprised of the following doses:
    - 4.5 mg dose - 80% of the participants
    - 3.0 mg dose - 15% of the participants
    - 1.5 mg dose - 5% of the participants

In addition, the planned sample size provides at least 95% power to detect a difference between treatment arms at Week 40 after randomization regardless of changes in background OAM, assuming that participants stay on treatment and do not take rescue antihyperglycemic medication. This sample size is based on a treatment difference of 0.60% with a common standard deviation of 1.1%, using a 2-group t-test and a 0.05 two-sided significance level. The treatment differences are based on the same assumptions stated above for the treatment-regimen estimand with a 15% discontinuation rate.

6. Supporting Documentation

6.1. Appendix 1: Demographic and Baseline Characteristics

A listing of patient demographics will be provided. All demographic and baseline clinical characteristics will be summarized by study treatment for the patients in the modified intent-to-treat (mITT) population. Baseline demographic and clinical characteristics of special interest include but not limited to: age, gender, race, ethnicity (applicable to only US population), weight, BMI, region, HbA1c, fasting glucose, duration of type 2 diabetes mellitus, and estimated glomerular filtration rate.

6.2. Appendix 2: Treatment Compliance

Treatment compliance for each visit interval is defined as taking at least 75% of the required doses of study drug. Similarly, a participant will be considered significantly noncompliant if he or she is judged by the investigator to have intentionally or repeatedly taken more than the prescribed amount of medication (more than 125%). Compliance over the study period will be calculated using the number of doses administered (regardless of the actual dose in mg administered) divided by the total number of doses expected to be administered ×100 over the study period. Treatment compliance will be summarized descriptively in the study period by treatment using the mITT population.

6.3. Appendix 3: Concomitant Therapy

The prespecified concomitant medications of interest will be summarized by treatment at randomization. Additionally, medications of interest initiated after randomization and change to medications of interest used at randomization will be summarized. The concomitant therapies will be mapped using the World Health Organization Drug dictionary in the clinical trial database.

The concomitant medications of interest include the following groups of medication:

- utilization of the following medications in study period
  - antihypertensive therapy
  - lipid lowering therapy
- rescue therapy due to severe persistent hyperglycemia
- initiation of antihyperglycemic medication after study treatment discontinuation
- initiation of the following medications in study period:
  - antidiarrheal medication
  - antiemetic medication

6.4. Appendix 4: Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

- Summary of AEs, provided as a dataset which will be converted to an XML file. Both SAEs and 'Other' non-SAEs are summarized: by treatment group, by MedDRA PT.
    - An AE is considered 'Serious' whether or not it is a TEAE.
    - An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
        - the number of participants at risk of an event
        - the number of participants who experienced each event term
        - the number of events experienced.
- For each SAE, these additional terms are provided for EudraCT:
    - the total number of occurrences causally related to treatment
    - the total number of deaths
    - the total number of deaths causally related to treatment.
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/subjects in every treatment group may be excluded if a 5% threshold is chosen. Allowable thresholds include 0% (all events), 1%, 2%, 3%, 4%, and 5%.
- AE reporting is consistent with other document disclosures for example, the CSR, manuscripts, and so forth.

Demographic table including the following age ranges required by EudraCT: in utero, preterm newborn infants (gestational age <37 weeks), newborns (0 to 27 days), infants and toddlers (28 days to 23 months), children (2 to 11 years), adolescents (12 to 17 years), adults (18 to 64 years), (65 to 85 years), and (85 years and over).

6.5. Appendix 5: Searching Criteria for Adverse Events of Special Interest

The AEs of special interest analyses are detailed in Section 4.6.2. The search criteria for each AESI are stored in CLUWE: T:\prd\ly3298176\common\AESI_Lab\Search criteria AESIs_TZP.xlsx.

7. References

Rubin DB. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, Inc.; 1987.

Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. *J Statl Plan Inference.* 2001;99(1): 25-40. https://doi.org/10.1016/S0378-3758(01)00077-5

Signature Page for VV-CLIN-056526 v1.0

Approval	PPD 01-Dec-2022 21:49:15 GMT+0000

Signature Page for VV-CLIN-056526 v1.0