Statistical Analysis Plan I8F-MC-GPHE (2)

A Randomized, Open-Label, Parallel-Group, Two-Arm, Phase 4 Study to Evaluate the Long-Term Efficacy and Safety of Tirzepatide Compared with Intensified Conventional Care in Adults When Initiating Treatment Early in the Course of Type 2 Diabetes (SURPASS-EARLY)

NCT05433584

Approval Date: 19-NOV-2025

Title Page

Protocol Title: A Randomized, Open-Label, Parallel-Group, Two-Arm, Phase 4 Study to Evaluate the Long-Term Efficacy and Safety of Tirzepatide Compared with Intensified Conventional Care in Adults When Initiating Treatment Early in the Course of Type 2 Diabetes (SURPASS-EARLY)

Protocol Number: I8F-MC-GPHE

Compound Number: LY3298176

Short Title: Efficacy and Safety of Tirzepatide Compared with Intensified Conventional Care in Adults when Treatment is Initiated Early for Type 2 Diabetes

Acronym: SURPASS-EARLY

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana, USA 46285

Regulatory Agency Identifier Number(s)

Registry	ID
IND	128801
EudraCT	2022-000130-42
EU trial number	2022-501073-40-00

Confidential Information

The information contained in this document is confidential and the information contained within it may not be reproduced or otherwise disseminated without the approval of Eli Lilly and Company or its subsidiaries

Note to Regulatory Authorities: This document may contain protected personal data and/or commercially confidential information exempt from public disclosure. Eli Lilly and Company requests consultation regarding release/redaction prior to any public release. In the United States, this document is subject to Freedom of Information Act (FOIA) Exemption 4 and may not be reproduced or otherwise disseminated without the written approval of Eli Lilly and Company or its subsidiaries.

Document ID: VV-CLIN-056527

Approved on 19 Nov 2025 GMT

Table of Contents

Title Page 1
Table of Contents 2
Version History 6
1. Introduction 7
1.1. Objectives, Endpoints, and Estimands 7
1.2. Study Design 14
2. Statistical Hypotheses 18
2.1. Multiplicity Adjustment 18
3. Analysis Sets 20
4. Statistical Analyses 21
4.1. General Considerations 21
4.1.1. Baseline Definition 22
4.1.2. Analysis Method 22
4.2. Participant Dispositions 28
4.3. Primary Endpoint Analysis 28
4.3.1. Primary Analysis Relative to the Treatment-Regimen Estimand 28
4.3.2. Primary Analysis Relative to the Efficacy Estimand 28
4.4. Secondary Endpoint Analysis 29
4.4.1. Key Secondary Endpoints 29
4.5. Additional Secondary 29
4.6. Exploratory Endpoints 32
4.6.1. Exploratory Analysis not Specified in the Protocol 35
4.7. Safety Analyses 37
4.7.1. Extent of Exposure 37
4.7.2. Adverse Events 37
4.7.3. Special Safety Topics 39
4.7.4. Vital Signs 43
4.7.5. Clinical Laboratory Evaluation 44
4.7.6. Device Product Complaints 44
4.8. Other Analyses 45
4.8.1. Subgroup Analyses 45
4.9. Interim Analyses 45
4.10. Changes to Protocol-Planned Analyses 46
5. Sample Size Determination 47
6. Supporting Documentation 48

6.1. Appendix 1: Demographic and Baseline Characteristics 48
6.2. Appendix 2: Treatment Compliance 48
6.3. Appendix 3: Concomitant Therapy 48
6.4. Appendix 4: Clinical Trial Registry 48
6.5. Appendix 5: Searching Criteria for Adverse Events of Special Interest 49
7. References 50

Approved on 19 Nov 2025 GMT

Table of Contents

Table **Page**

Table GPHE.4.1 Imputation Procedure 25
Table GPHE.4.2 Study Discontinuation Reasons....26
Table GPHE.4.3 Additional Secondary Efficacy Analyses not Controlled for Type I Error..30
Table GPHE.4.4 Exploratory Efficacy Analyses Not Controlled for Type I Error....33
Table GPHE.4.5 Non- protocol Specified Exploratory Efficacy Analyses Not Controlled for Type I Error....36
Table GPHE.4.6 Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurements....44

Table of Contents

Figure **Page**

Figure GPHE.1.1. Illustration of study design for clinical protocol study I8F-MC-GPHE.....15

Figure GPHE.2.1. Type I error control strategy for primary and key secondary efficacy endpoints. 19

Figure GPHE.4.1. Treatment journey of a trial participant in relation to permanent discontinuation of study intervention or study. 24

Version History

This Statistical Analysis Plan (SAP) for Study GPHE is based on Protocol I8F-MC-GPHE(a) dated 20 June 2022. SAP version 1 was approved prior to First Patient Visit (FPV).

SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	27 July 2022	Not Applicable	Original version
2	See date on Page 1	Added additional details regarding the imputation methods.	Clarified the distinction of imputing discontinuation strategy
		Updated the analysis language and removed the repetitive texts	Clarification
		Added additional prespecified exploratory endpoints	Potential plan for prespecified CVD risk analysis
		Updated Region category	To be consistent with protocol
		Minor update in endpoints definition	For consistency with the data collection procedures
		Updated the Analysis set	To align with new strategies
		Throughout the SAP, Minor changes and reorganize	For clarification, no change to analysis methodologies, so not details
		Throughout SAP, Changes LS means to model-based estimate means	Update language to better align with unconditional average treatment effect estimate

Approved on 19 Nov 2025 GMT

1. Introduction

Tirzepatide is a once weekly glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (GLP-1) receptor agonist that combines gastric inhibitory polypeptide receptor (GIPR) agonism and glucagon-like peptide-1 receptor agonist (GLP-1R) agonism into a single molecule developed for the treatment of type 2 diabetes (T2D). In a series of 5 global Phase 3 SURPASS studies, tirzepatide demonstrated clinically meaningful reductions in hemoglobin A1c (HbA1c) and body weight, which were greater than placebo, semaglutide 1 mg, titrated insulin degludec, and titrated insulin glargine (Rosenstock et al. 2021; Frías et al. 2021; Ludvik et al. 2021; Del Prato et al. 2021; Dahl et al. 2022). HbA1c results of Phase 3 SURPASS studies show that tirzepatide may potentially address an important clinical care gap in the treatment of T2D. This study is designed to determine long-term efficacy and safety of tirzepatide 15 mg or maximum tolerated dose (MTD) compared with intensified conventional care in adults with a 4 years of T2D since diagnosis. This study aims to test the hypothesis that, during the course of T2D, when compared to other existing treatment options, early initiation of tirzepatide treatment may improve glycemic control and lower body weight during the first 2 years. Moreover, these improvements may be maintained through 4 years without the need to initiate additional glucose-lowering therapy as well as delay the deterioration of β-cell function. Existing treatment options are represented in this trial by the intensified conventional care. Treatments in the intensified conventional care arm include a range of glucose-lowering medications used in clinical practice and supported by treatment guidelines.

This SAP describes the pre-specified statistical analyses for Study GPHE. These analyses apply to efficacy and safety data.

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Primary	
To demonstrate that tirzepatide is non-inferior to intensified conventional care in hemoglobin A1c (HbA1c) change from baseline to Week 104 in participants with type-2 diabetes (T2D)	• Change in HbA1c from baseline to Week 104
Key Secondary	
To demonstrate that tirzepatide is superior to intensified conventional care at Week 104	• Change in HbA1c from baseline to Week 104 • Change in weight from baseline to Week 104 • Change in waist circumference from baseline to Week 104
To demonstrate that tirzepatide is non-inferior to intensified conventional care at Week 208	• Change in HbA1c from baseline to Week 208
To demonstrate that tirzepatide is superior to intensified conventional care at Week 208	• Change in HbA1c from baseline to Week 208 • Change in weight from baseline to Week 208 • Change in waist circumference from baseline to Week 208

Objectives	Endpoints
Additional Secondary	
To assess the clinical effect of tirzepatide compared to intensified conventional care at Week 104	• Change in FSG (via central lab) from baseline to Week 104 • Change in lipids from baseline (total cholesterol, HDL, LDL, VLDL, and TG) to Week 104 • HbA1c <5.7%, ≤6.5%, and <7% at Week 104 • Weight loss from baseline of ≥5%, ≥10%, and ≥15% at Week 104 • A composite endpoint ("HbA1c ≤6.5% & Weight Loss ≥10% & No-Hypoglycemia") at Week 104 defined as o HbA1c ≤6.5% o weight loss ≥10%, and o no level 2 or severe hypoglycemia, defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia • Change in β-cell function and insulin sensitivity from baseline to Week 104 assessed in OGTT o β-cell glucose sensitivity based on the modeling analysis using Mari's model[a] o insulin secretion rate at a fixed glucose concentration (adjusted and unadjusted for basal potentiation) from modeling analysis using Mari's model[a] o postprandial insulin sensitivity indices ▪ Matsuda index ▪ OGIS o fasting and post-challenge glucose (total and incremental $AUC_{0\text{-}180min}$) o fasting and post-challenge insulin (total and incremental $AUC_{0\text{-}180min}$) • Changes in Impact of Weight on Quality of Life-Lite Clinical Trials Version (IWQOL-Lite-CT) – Physical-Functioning domain from baseline to Week 104

Approved on 19 Nov 2025 GMT

Objectives	Endpoints

To assess the clinical effect of tirzepatide compared to intensified conventional care at Week 208	• Change in FSG (via central lab) from baseline to Week 208
	• Change in lipids from baseline (total cholesterol, HDL, LDL, VLDL, and TG) to Week 208
	• HbA1c <5.7%, ≤6.5%, and <7% at Week 208
	• Weight loss from baseline of ≥5%, ≥10%, and ≥15% at Week 208
	• A composite endpoint ("HbA1c ≤6.5% & Weight Loss ≥10% & No-Hypoglycemia") at Week 208, defined as
	  o HbA1c ≤6.5%
	  o weight loss ≥10%, and
	  o no level 2 or severe hypoglycemia, defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia
	• Change in β-cell function and insulin sensitivity from baseline to Week 208 of the following parameters assessed in OGTT:
	  o β-cell glucose sensitivity based on modeling analysis using Mari's model[a]
	  o insulin secretion rate at a fixed glucose concentration (adjusted and unadjusted for basal potentiation) from modeling analysis using Mari's model[a]
	  o postprandial insulin sensitivity indices:
	  ▪ Matsuda index
	  ▪ OGIS
	  o fasting and post-challenge glucose (total and incremental $AUC_{0\text{-}180min}$)
	  o fasting and post-challenge insulin (total and incremental $AUC_{0\text{-}180min}$)
	• Changes in IWQOL-Lite-CT – Physical-Functioning domain from baseline to Week 208

Approved on 19 Nov 2025 GMT

Objectives	Endpoints
Exploratory	

To assess the clinical effect of tirzepatide compared to intensified conventional care at Week 104	• Change in daily average 7-point self-monitored blood glucose profiles
	• Change in BMI from baseline
	• Change in urinary albumin to creatinine ratio from baseline
	• Change in eGFR from baseline
	• Change in number of antihyperglycemic medications from baseline
	• Time to treatment failure, defined as HbA1c ≥7% confirmed at 2 consecutive study visits after Week 32
	• The change from baseline in the following parameters assessed during OGTT:
	  ○ glucagon concentration at fasting and post-challenge (total and incremental $AUC_{0\text{-}180min}$)
	  ○ insulin/glucagon ratios at fasting and post-challenge (total and incremental $AUC_{0\text{-}180min}$)
	  ○ C-peptide concentration at fasting and post-challenge (total and incremental $AUC_{0\text{-}180min}$)
	  ○ insulinogenic index
	  ○ insulin clearance
	• Changes in SF-36 v2 acute form physical functioning domain score from baseline
	• Changes in IWQOL-Lite-CT (Total, Physical, and Psychosocial) score from baseline
	• Ability to Perform Physical Activities of Daily Living (APPADL)
	• Impact of Weight on Self Perception (IW-SP)

Approved on 19 Nov 2025 GMT

Objectives	Endpoints
To assess the effect of tirzepatide compared to intensified conventional care at Week 208	• Change in daily average 7-point self-monitored blood glucose profiles • Change in BMI from baseline • Change in urinary albumin to creatinine ratio from baseline • Change in eGFR from baseline • Change in number of antihyperglycemic medications from baseline • Time-averaged HbA1c over the entire study period • Time to treatment failure, defined as HbA1c ≥7% confirmed at 2 consecutive study visits after Week 32 • The change from baseline in the following parameters assessed during OGTT: o glucagon concentration at fasting and post-challenge (total and incremental $AUC_{0\text{-}180min}$) o insulin/glucagon ratios at fasting and post-challenge (total and incremental $AUC_{0\text{-}180min}$) o C-peptide concentration at fasting and post-challenge (total and incremental $AUC_{0\text{-}180min}$) o insulinogenic index o insulin clearance • Changes in SF-36 v2 acute form physical functioning domain score from baseline • Changes in IWQOL-Lite-CT (Total, Physical, and Psychosocial) score from baseline • Ability to Perform Physical Activities of Daily Living (APPADL) • Impact of Weight on Self Perception (IW-SP)

Approved on 19 Nov 2025 GMT

Objectives	Endpoints
Exploratory (not specified in protocol)	
To assess the effect of tirzepatide compared to intensified conventional care at Week 104 and at Week 208	• The change from baseline in the following parameters assessed during OGTT (calculating with C-peptide or insulin): o HOMA-S, HOMA-IR, HOMA-B o The following composite endpoints where Hypoglycemia is defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia: o HbA1c <5.7% & No-Weight Gain & No-Hypoglycemia o HbA1c <5.7% & Weight Loss ≥5% & No-Hypoglycemia o HbA1c <5.7% & Weight Loss ≥10% & No-Hypoglycemia o HbA1c <5.7% & Weight Loss ≥15% & No-Hypoglycemia o HbA1c <5.7% & Weight Loss ≥20% & No-Hypoglycemia o HbA1c ≤6.5% & No-Weight Gain & No-Hypoglycemia o HbA1c ≤6.5% & Weight Loss ≥5% & No-Hypoglycemia o HbA1c ≤6.5% & Weight Loss ≥15% & No-Hypoglycemia o HbA1c ≤6.5% & Weight Loss ≥20% & No-Hypoglycemia o HbA1c <7.0% & No-Weight Gain & No-Hypoglycemia o HbA1c <7.0% & Weight Loss ≥5% & No-Hypoglycemia o HbA1c <7.0% & Weight Loss ≥10% & No-Hypoglycemia o HbA1c <7.0% & Weight Loss ≥15% & No-Hypoglycemia o HbA1c <7.0% & Weight Loss ≥20% & No-Hypoglycemia • Change from baseline in high sensitivity (hsCRP) and microalbumin • Number of antihypertensive medications • Number of lipid-lowering medications
To assess the change in predicted 10-year cardiovascular risk of tirzepatide compared to intensified conventional care at Week 104 and at Week 208 from baseline	• change in 10-year cardiovascular risk • percent change in 10-year cardiovascular risk

Approved on 19 Nov 2025 GMT

Objectives	Endpoints
To assess the effect of tirzepatide compared to intensified conventional care at Week 208	• HbA1c <5.7%, ≤6.5%, and <7% in participants achieving HbA1c <5.7%, ≤6.5%, and <7% respectively, at Week 104 • Weight loss of ≥5%, ≥10%, and ≥15% in participants achieving Weight loss of ≥5%, ≥10%, and ≥15% respectively, at Week 104 • HbA1c ≤6.5% & Weight Loss ≥10% & No-Hypoglycemia in patients achieving HbA1c ≤6.5% & Weight Loss ≥10% & No-Hypoglycemia at Week 104

Abbreviations: AUC = area under the curve; BG = blood glucose; FSG = fasting serum glucose; HDL = high-density lipoprotein; LDL = low-density lipoprotein; OGIS = oral glucose insulin sensitivity; OGTT = oral glucose tolerance test; TG = triglyceride; VLDL = very-low-density lipoprotein.

[a] Mari et al. 2002.

Note: Primary and key secondary endpoints are controlled for multiplicity.

Primary estimand

There will be 2 primary estimands evaluated in this study, for evaluating the primary and key secondary objectives of GPHE. The details of the plan for each estimand can be found in Section 2.1:

- The treatment-regimen estimand is accepted by the FDA and has been used to inform the US Prescribing Information and may be used for other regions as well. This estimand aims at reflecting how participants with T2D are treated in clinical practice and considers 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 antihyperglycemic medication for severe, persistent hyperglycemia. This estimand will be used in publications to inform prescribers/physicians and may be submitted to other regulatory agencies.

Treatment-regimen estimand

The treatment regimen estimand serves as a primary estimand for this trial. The treatment-regimen estimand answers the following question of interest for the primary objective: What is the treatment difference in change from baseline in HbA1c after 104 weeks of treatment in participants with T2D regardless of treatment discontinuation for any reason and regardless of initiation of antihyperglycemic rescue treatment or prohibited medication (except for non-study tirzepatide) or change in the background metformin treatment?

The treatment-regimen estimand is described by the following attributes.

- *Population*: Participants with T2D inadequately controlled with diet and exercise with a background of metformin at a stable dose. Further details can be found in Section 5 of the study protocol.

- *Endpoint*: Change from baseline to Week 104 after randomization in HbA1c.
- *Treatment condition*: The randomized treatment regardless of adherence to treatment with or without antihyperglycemic rescue medication or change in background metformin medication. Further details on study treatment and concomitant therapy, including rescue, treatments can be found in Section 6 of the study protocol.
- *Intercurrent events*: "treatment discontinuation for any reason" and "initiation of rescue antihyperglycemic treatment or prohibited medication (except for non-study tirzepatide) or change in background metformin medication" are addressed by the treatment condition. There are no remaining intercurrent events.
- *Population-level summary*: Difference in mean changes between treatment conditions.

Efficacy estimand

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 104 after randomization in participants with T2D regardless of changes in background metformin medication, assuming that participants had stayed on treatment and not taken antihyperglycemic rescue medication?

The efficacy estimand is described by the following attributes.

- *Population*: Participants with T2D inadequately controlled with diet and exercise with a background of metformin. Further details can be found in Section 5 of the study protocol.
- *Endpoint*: HbA1c change from baseline to Week 104 after randomization.
- *Treatment condition*: The randomized treatment, regardless of changes to the background metformin medication. Further details on study treatment can be found in Section 6 of the study protocol.
- *Intercurrent event*: "Treatment discontinuation for any reason" and "Initiation of antihyperglycemic rescue treatment or prohibited medication" will be addressed using the following strategies:
    - Had participants stayed on treatment (hypothetical strategy).
    - Had participants not taken antihyperglycemic rescue medication or prohibited medication (hypothetical strategy).
- *Population-level summar*y: Difference in mean changes between treatment conditions.

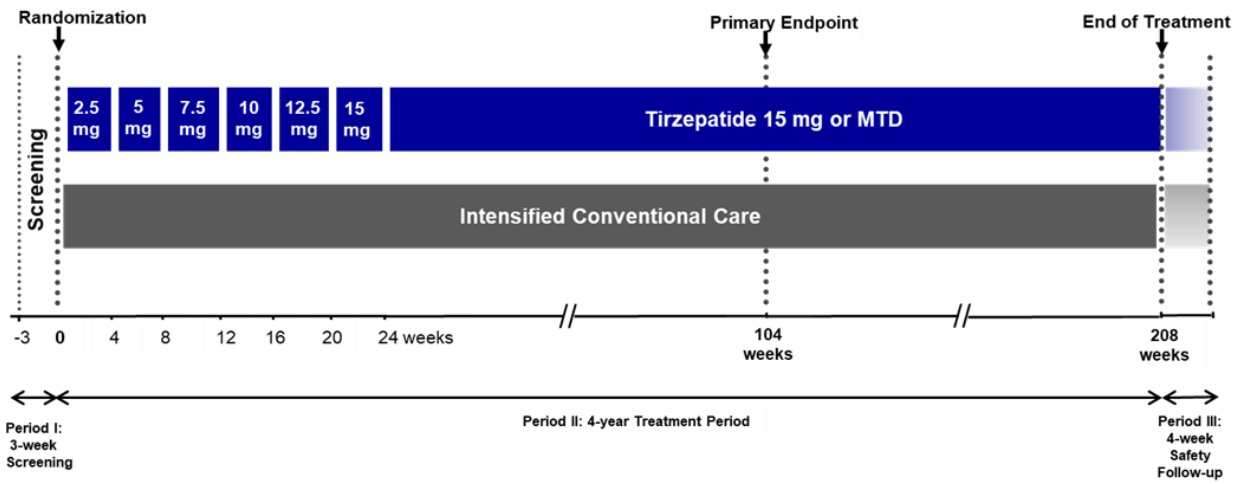
Both the efficacy and treatment-regimen estimands will be evaluated for the primary and all key secondary objectives. The population, treatment condition, intercurrent events, and population-level summary specified above for each estimand for the primary objective will also apply to the key secondary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

1.2. Study Design

SURPASS-EARLY is a randomized, open-label, parallel-group, two-arm, multicenter, multinational, Phase 4 study to investigate the long-term efficacy and safety of tirzepatide 15 mg or MTD compared to intensified conventional care in participants with up to 4 years of T2D history. Treatments in the intensified conventional care arm include a range of glucose-lowering medications used in clinical practice and supported by treatment guidelines. They will be used in

accordance with the product labels approved locally. This study will enroll participants with T2D inadequately controlled with diet, exercise, and metformin. The primary endpoint will be the mean change in HbA1c from baseline to 104 weeks.

This study includes an approximate 3-week screening period, a 208-week treatment period, and a 4-week safety follow-up (Figure GPHE.1.1).



Abbreviation: MTD = maximum tolerated dose.

Figure GPHE.1.1. Illustration of study design for clinical protocol study I8F-MC-GPHE.

Period I: Screening

Visit 1

Interested individuals will sign the appropriate informed consent document(s) prior to initiating any procedure.

The investigator will review medical history and other inclusion and exclusion criteria prior to any diagnostic procedures. If the participant is eligible after this review, then the site will perform the diagnostic procedures to confirm eligibility.

Visit 2

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

Participants arrive to the clinic in the fasting state as specified in Section 10.8, Appendix 8 of the study protocol.

For participants meeting eligibility requirements, an OGTT test to assess β-cell function will be performed at Visit 2.

A dilated fundoscopic examination will be scheduled and completed between Visit 2 and Visit 3.

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

At Visit 3 participants who are confirmed enrollment criteria will be randomly assigned to an intervention group.

First tirzepatide dose

Study personnel will observe as the participant injects the first dose of tirzepatide. The date and time of the first dose of tirzepatide should be recorded on the case report form (CRF).

First intensified conventional care dose

Study personnel will provide appropriate instruction for treatment with the antihyperglycemic medication selected as the study intervention. Additional training may be provided at subsequent visits, as needed.

Study personnel will provide product-specific injectable "instructions for use" information to the participants upon request.

Visit 4 through Visit 24 (Week 4 through Week 200) Treatment period

Participants complete all visit procedures.

Visit 25 (Week 208) or ED visit

Participants complete all visit procedures.

The investigator will determine a participant's transition from study treatment to another diabetes treatment.

Period III: Safety follow-up visit

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

Patients discontinuing the study early will also be asked to perform the safety follow-up visit, so that the safety follow-up visit will be their final visit. During the safety follow-up period, patients will not receive study drug. Initiation of new antihyperglycemic therapy for the safety follow-up period will not be classified as "rescue therapy."

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 computed-generated random sequence using an interactive web-response system (IWRS).

Participants will be randomly assigned in a 1:1 ratio to tirzepatide or intensified conventional care, based on these stratification factors at screening:

- region: North America (NA), Latin America (LATAM), Europe + Israel (EU + Israel)
- age (≤40 and ≥41 years), and
- HbA1c (≤8.0% or >8.0%)

2. Statistical Hypotheses

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

- **$H_{1,0}$**: tirzepatide 15 mg or MTD is inferior to intensified conventional care with respect to HbA1c mean change from baseline at Week 104 by 0.3% or more.

The null hypotheses corresponding to the key secondary objectives are as follows:

- **$H_{2,0}$**: tirzepatide 15 or MTD is not superior to intensified conventional care with respect to HbA1c mean change from baseline at Week 104.
- **$H_{3,0}$:** tirzepatide 15 or MTD is not superior to intensified conventional care with respect to mean weight change from baseline at Week 104.
- **$H_{4,0}$:** tirzepatide 15 or MTD is not superior to intensified conventional care with respect to mean waist circumference change from baseline at Week 104.
- **$H_{5,0}$:** tirzepatide 15 mg or MTD is inferior to intensified conventional care with respect to HbA1c mean change from baseline at Week 208 by 0.3% or more.
- **$H_{6,0}$:** tirzepatide 15 mg or MTD is not superior to intensified conventional care with respect to HbA1c mean change from baseline at Week 208.
- **$H_{7,0}$:** tirzepatide is not superior to intensified conventional care with respect to weight mean change from baseline at Week 208.
- **$H_{8,0}$:** tirzepatide is not superior to intensified conventional care with respect to waist circumference mean change from baseline at Week 208.

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.

A prespecified graphical scheme (Bretz et al. 2009, 2011) shown in Figure GPHE.2.1 below will be implemented to control the family-wise error rate at a 1-sided alpha level of 0.025 for testing the null hypotheses stated in Section 2. More specifically, multiple testing adjusted p-values described by Bretz et al. (2009) will be calculated, and any hypothesis tests with a multiple testing adjusted 1-sided p-value of less than 0.025 will be considered statistically significant. This graphical approach is a closed testing procedure; hence, it strongly controls the family-wise error rate across all endpoints (Bretz et al. 2009, 2011; Alosh et al. 2014).

Unless otherwise specified, there will be no adjustment for multiple comparisons for any other analyses outside the primary and key secondary endpoints. The testing scheme will be finalized before the first participant visit.

For analysis within each estimand, type 1 error rate control strategy for evaluation of primary and key secondary objectives is illustrated in Figure GPHE.2.1.

Approved on 19 Nov 2025 GMT

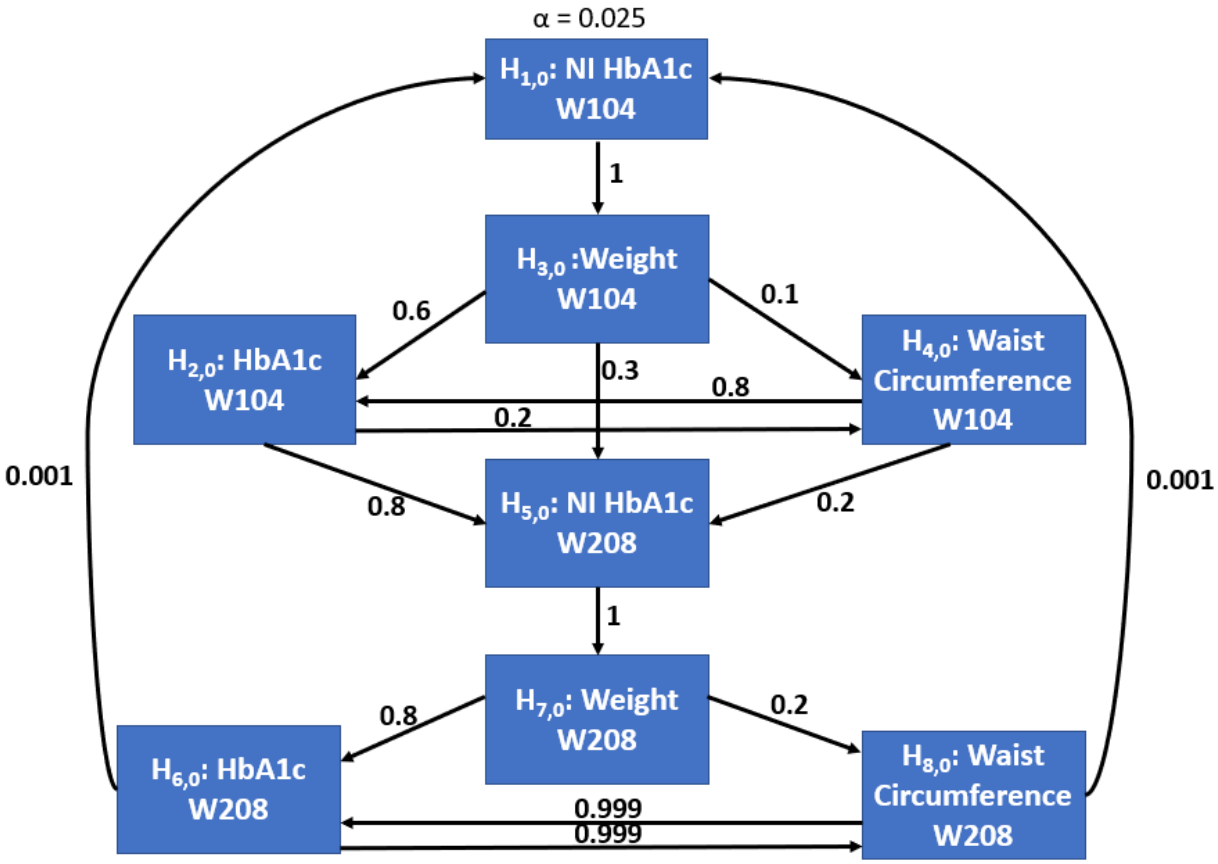


Figure GPHE.2.1. Type I error control strategy for primary and key secondary efficacy endpoints.

3. Analysis Sets

This table defines the analysis population 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 group.
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: 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 randomized population, excluding data after permanent discontinuation of treatment or initiation of antihyperglycemic rescue medication or prohibited medication.
Full Analysis Set: 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 randomized population, regardless of adherence to treatment or initiation of antihyperglycemic rescue medication or prohibited medication (excluding non-study Tirzepatide). Data after taking non-study tirzepatide will be excluded from analysis.
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 antihyperglycemic rescue antihyperglycemic medication.

4. Statistical Analyses

4.1. General Considerations

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly) or its designee.

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 statistical analysis plan or clinical study report. Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (for example, 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.

Primary analysis will be performed when all participants complete Week 104 or discontinue treatment. Final analysis will be performed after Last Participant Last Visit (LPLV).

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 will be given at a 2sided [95%] level. In statistical summaries and analyses, all data will be analyzed by randomized treatment assignment. Participants will be analyzed according to the treatment they were randomly assigned to, regardless of the treatment actually received.

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. Selected safety analyses may be conducted after excluding data on rescue therapy or data after starting another antihyperglycemic medication.

The ‘treatment-regimen’ estimand will be assessed for primary and key secondary and selected additional secondary efficacy objectives using the FAS for the purpose of informing promotional materials. For publications and other purposes, the assessment of efficacy objectives will be guided by the “efficacy” estimand using the EAS.

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 (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 Kaplan-Meier method may 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. Fisher’s exact test will be used to examine the

treatment difference in categorical outcomes. Logistic regression may be used to examine the treatment difference in binary efficacy outcomes. The negative binomial regression model will be used for the treatment comparison of discrete count measures if deemed appropriate.

4.1.1. Baseline Definition

Baseline is defined as the last non-missing measurement recorded on or before the randomization visit, prior to first dose of treatment. If there are no doses of study intervention, the baseline will be defined as the last available non-missing measure on or prior to randomization. For laboratory values, baseline needs to be prior to or within one hour after the first dose time. In cases where the measurement is taken on the same day (where the time is not collected or not reliable) as the first dose, this measurement will be used as the baseline value for data analysis. For patient reported outcomes, data collected at Visit 3, regardless of the timing relative to first dose, will serve as baseline. For the safety related parameters, baseline is defined as the start of screening and ends prior to the first dose of study intervention (typically at Week 0).

4.1.2. Analysis Method

4.1.2.1. Analysis Method for Treatment Regimen Estimand

4.1.2.1.1. Method for Continuous Variables

For the assessment of continuous efficacy measures guided by the treatment regimen estimand, data for participants with missing values at final timepoint visit will be imputed based on methods described in Section 4.1.2.1.3, and then analyzed using an ANCOVA with a variance estimator that is robust to model misspecification and heteroscedasticity (Ye et al. 2022, FDA 2023). The model will include

- treatment group as a factor variable
- region as a factor variable
- age
- a single strata variable for stratification factors (HbA1c at baseline [≤8.0%, >8.0%]
- baseline value (of the dependent variable)

For analyses of HbA1c, the baseline HbA1c category will be removed. Statistical comparisons will be performed between tirzepatide MTD and intensified conventional care arm. The final inference will be derived using Rubin's rule which combines estimates from multiple imputed datasets. The imputation procedure for creating a single imputed dataset is detailed in Table GPHE.4.1.

Some parameters may be log-transformed before fitting the ANCOVA as specified in Section 4.4, Section 4.5, and Section 4.6, with the associated log-transformed baseline value as a covariate. In these cases, MBE and 95% CIs for each treatment group and treatment difference will be back-transformed and presented as mean percent change from baseline and as relative treatment difference to intensified conventional care in percent change.

4.1.2.1.2. Method for Binary Variables

For binary efficacy measures derived from a continuous variable and guided by the treatment regimen estimand, the missing value in the underlying continuous variable will be imputed first

based on methods described in Section 4.1.2.1.3 and then the corresponding binary variable will be derived.

For the assessment of binary efficacy measures, a logistic regression model will be used to examine the treatment difference with missing values at final timepoint imputed based on methods described in Section 4.1.2.1.3. The model will include:

- treatment group as a factor variable
- region as a factor variable
- age
- a single strata variable for stratification factors (HbA1c at baseline [≤8.0%, >8.0%]
- baseline value (of the dependent variable)

For analyses of HbA1c, the baseline HbA1c category will be replaced with continuous baseline HbA1c measurements. The unconditional treatment group effect will be assessed by risk difference and relative risk based on the delta-method using the formula provided (Ye et al. 2023). The final inference will be derived using Rubin's rule by combining estimates from multiple imputed datasets, with the imputation procedure for creating a single imputed dataset detailed in Table GPHE.4.1. The estimated treatment group-specific risk, risk difference, relative risk, p-value and 95% CI will be presented.

4.1.2.1.3. Primary Multiple Imputation Strategy

Discontinuation is expected to be uncommon. Participants who discontinue study intervention (that is, discontinue study treatment) will be encouraged to continue in Study GPHE for the treatment period and follow-up period (note, the case report forms [CRFs] use "phase," which is interchangeable with the "period" in SAP). In this section, study discontinuation refers to treatment phase discontinuation captured in CRF.

If there are occurrences of missing data despite all reasonable precautions, missing data will be imputed in a manner consistent with what the values would likely have been, had they been collected.

In general, 5 scenarios are considered regarding the treatment journey of a trial participant relative to the primary time point shown in Figure GPHE.4.1.

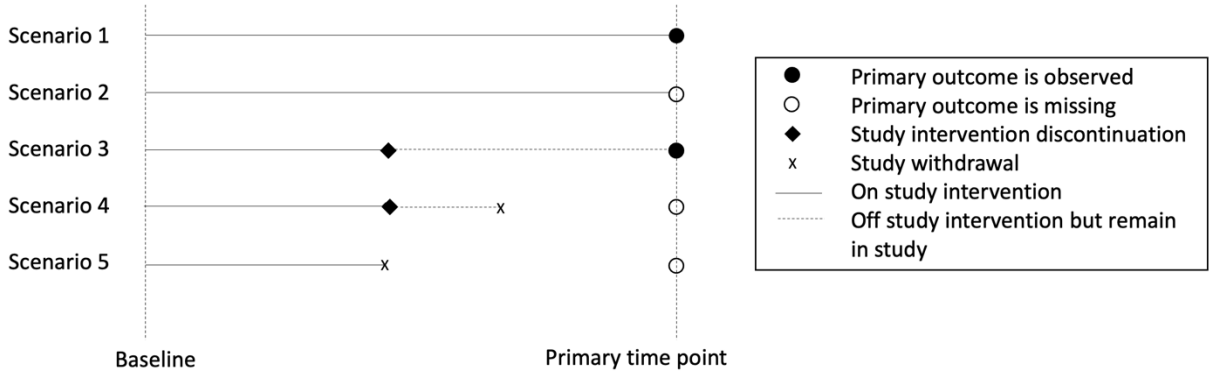


Figure GPHE.4.1. Treatment journey of a trial participant in relation to permanent discontinuation of study intervention or study.

Scenario 5 also covers situations where a participant discontinues from the study intervention and then comes back for a study discontinuation visit for the same discontinuation reason on or prior to a scheduled visit.

In principle, missing data due to permanent discontinuation of the study intervention (Scenario 4 and 5A) will be imputed by treatment group using retrieved dropouts (MI-RD), namely using multiple imputation based on data retrieved from participants who permanently discontinued the study intervention but continued in Study GPHE with non-missing measurements from the same treatment group. For scenarios 4 and 5A, if there are not enough retrieved dropouts to provide a reliable imputation model (that is, the model implemented does not converge), an alternative multiple imputation method (Return to Baseline method) will be used. Any missing values at baseline will be imputed in the same model when imputing the endpoint.

Table GPHE.4.1 Imputation Procedure

Scenario (Study Intervention/Study Discontinuation; Missingness at Endpoint)	Methods to Handle Missing Values at Endpoint
1. No study intervention discontinuation; no missing value	N/A
2. No study intervention discontinuation; with missing value	Use observed data, including baseline and all post-baseline visits, across all time points from Scenario 1 and 2 to impute under the MAR assumption by treatment group. There should be very few such cases in a clinical trial. No additional covariates will be included in the imputation model.
3. Study intervention discontinuation; no missing value	N/A
4. Study intervention discontinuation with study discontinuation at a subsequent visit or completion of treatment period; with missing value	Missing values at the endpoint visit will be imputed through MCMC (predictive mean matching method), using baseline and endpoint visit data from Scenario 3 (MI-RD) and impute missing values by treatment group. No additional covariates will be included in the imputation model.
5. Study discontinuation resulting in study intervention discontinuation; with missing value	5A: If the study discontinuation is possibly related to study intervention (see study discontinuation reasons classified as Category 5A in Table GPHE.4.4), missing values at the endpoint visit will be imputed in the same way as Scenario 4. 5B: If the study discontinuation is clearly due to administrative reasons not related to study intervention (see study discontinuation reasons classified as Category 5B in Table GPHE.4.4), missing values at the endpoint visit will be imputed in the same way as Scenario 2.

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

Table GPHE.4.2 Study Discontinuation Reasons

Disposition Reason	Associated Sub-categories	Category
Adverse event		5A
Death		5A
Protocol deviation [a]		5A
Pregnancy		5A
Non-compliance with study drug		5A
Lack of efficacy		5A
Withdrawal by subject	Concern about study procedures/perceived risks	5A
	Scheduling conflicts	5A
	Subject is moving or has moved	5A
	Personal issue unrelated to trial	5A
	Due to epidemic/pandemic	5B
	Other (option to include a specify field)	5A
Physician decision	Concern about study procedures/perceived risks	5A
	Due to epidemic/pandemic	5B
	Other (option to include a specify field)	5A
Study terminated by sponsor		5B
Site terminated by sponsor		5B
Study terminated by IRB/ERB		5B
Lost to follow-up		5A
Other		5A

Abbreviations: ERB = ethical review board; IRB = institutional review board.

Note: For participants that discontinued the study due to emergency unblinding, the study disposition will be classified based upon the reason leading to emergency unblinding (for example, Adverse Event [5A]).

4.1.2.2. Analysis Method for Efficacy Estimand

4.1.2.2.1. Method for Continuous Variables

For the assessment of continuous efficacy measures guided by the efficacy estimand, a restricted maximum likelihood-based MMRM (Wang and Du 2023) analysis will be used. All the longitudinal observations at each scheduled postbaseline visit will be included in the analysis. The MMRM model will include these terms:

- treatment group as a factor variable
- visit as a factor variable
- region as a factor variable
- age
- a single strata variable for stratification factors (HbA1c at baseline [≤8.0%, >8.0%)]

- baseline value (of the dependent variable), and
- interaction between treatment group, visit.

For analyses of HbA1c, the baseline HbA1c category will be replaced with continuous baseline HbA1c measurements. The estimated treatment group effect and comparison between tirzepatide MTD and intensified conventional care at the scheduled visits will be reported together with the variability. The associated 2-sided 95% CI and corresponding p-values will also be reported. The sandwich estimator (Diggle et al. 1994) for the variance-covariance matrix will be used. An unstructured covariance matrix by treatment groups will be used to model the within-participant errors, assuming heteroscedasticity (variant covariance structure across treatment groups) and the measurements for different participants are independent.

If the model still fails to converge, the following covariance structures will be tested in this order for the simplified model:

- unstructured covariance matrix
- heterogeneous Toeplitz
- heterogeneous autoregressive
- heterogeneous compound symmetry
- homogeneous Toeplitz
- homogeneous autoregressive, and
- homogeneous compound symmetry.

The first covariance structure that converges will be used.

Some parameters may be log-transformed before fitting the MMRM as specified in Section 4.4, Section 4.5, and Section 4.6. Some parameters may be log-transformed before fitting the MMRM as specified in Section 4.4, Section 4.5, and Section 4.6. In these cases, model-based estimates and 95% CIs for each treatment group and treatment difference will be back-transformed and presented as mean percent change from baseline and as relative treatment difference to placebo in percent change.

If the data does not warrant the use of an MMRM model, then an ANCOVA model will be conducted, as described in Section 4.1.2.1.1, without the primary imputation method.

4.1.2.2.2. Method for Binary Variables

For binary efficacy measures derived from a continuous variable and guided by the efficacy estimand, the missing value in the underlying continuous variable will be imputed first based on methods described in Section 4.1.2.2.3 and then the corresponding binary variable will be derived.

For the assessment of binary efficacy objectives, the logistic regression model described in Section 4.1.2.1.2 will be used after imputation and estimation of treatment group specific risk, risk difference, relative risk, p-value, and 95% CI will be presented.

4.1.2.2.3. Handling of Missing Data

Missing data should be minimized at the best precaution. For continuous efficacy analyses guided by efficacy estimand, the hypothetical strategy will be used to handle the ICEs. For the MMRM analysis, only data collected before the occurrence of any ICEs will be used. Through the MMRM, the potential efficacy measures (after the ICEs) will be implicitly imputed as if

participants did not have ICEs. For the ANCOVA analysis, missing values at the last visit will be imputed through multiple imputation using all non-missing data (excluding data collected after ICEs) from the same treatment group under the missing at random (MAR) assumption. The same method will also be used to impute the missing value in the underlying continuous variables of binary outcomes.

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 104 and 208-week visit will be provided by study treatment. Reasons for prematurely discontinuing study treatment and study prior to the 104 and 208-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 104-week visit (Visit 16). 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 (% and mmol/mol) from baseline (postbaseline – 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 is 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 104 -week visit with the aid of an ANCOVA model and specified in Section 4.1.2.1.1.The response variable will be the primary measure and model terms will include treatment and region as fixed effects and baseline HbA1c and age as covariates. The primary analysis relative to the treatment-regimen estimand will be conducted utilizing HbA1c data in the FAS at baseline and at the 104 -week visit with the aid of an ANCOVA model and specified in Section 4.1.2.1.1.The response variable will be the primary measure and model terms will include treatment and region as fixed effects and baseline HbA1c and age as covariates.

If the upper limit of the CI is ≤0.3%, then tirzepatide 15 mg or MTD will be declared noninferior to intensified conventional care for change in HbA1c from baseline at Week 104.

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 104 -week visit with the aid of a mixed-model repeated measure (MMRM) and specified in Section 4.1.2.2.1 (MMRM).

The independent variables of the MMRM model are treatment, visit, treatment-by-visit interaction, and region as fixed effects, and age and baseline HbA1c as covariates. If the upper limit of the CI is ≤0.3%, then tirzepatide 15 mg or MTD will be declared noninferior to intensified conventional care to change in HbA1c from baseline.

4.4. Secondary Endpoint Analysis

4.4.1. Key Secondary Endpoints

The endpoints corresponding to secondary study objectives subject to Type I error rate control are specified in Section 1.1 under "Key Secondary (Controlled for Type I error) endpoints".

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

Key secondary objectives 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 change in HbA1c from Baseline at the 104- and 208-week Visit

Assessment of superiority of mean changes in HbA1c of tirzepatide 15 mg or MTD compared with intensified conventional care at Week 104 and Week 208 will be conducted using the same statistical models as those used for evaluating the primary objective in Section 4.3.1 (treatment-regimen estimand) and Section 4.3.2 (efficacy estimand). Inferences will be made based on the multiplicity testing framework outlined in Section 2.1.

Similarly, assessment of noninferiority of mean changes in HbA1C of tirzepatide 15 mg or MTD to intensified conventional care at Week 208 will be conducted in a manner similar to the primary analysis in Section 4.3.

4.4.1.2. Mean Change in Body Weight from Baseline at the 104- and 208-week Visit

The analysis for change in body weight from baseline (postbaseline – baseline) at Week 104 and Week 208 will be conducted in a manner similar to the primary analysis and specified in Section 4.1.2.1 (treatment-regimen estimand) and Section 4.1.2.2 (efficacy estimand).

4.4.1.3. Mean Change in Waist Circumference from Baseline at the 104- and 208-week Visit

The analysis for change in Waist circumference from baseline (postbaseline - baseline) at Week 104 and Week 208 will be conducted in a manner similar to the primary analysis and specified in Section 4.1.2.1 (treatment-regimen estimand) and Section 4.1.2.2 (efficacy estimand).

4.5. Additional Secondary

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

- Change in FSG (via central lab) from baseline
- HbA1c <5.7%, ≤6.5%, and <7%
- Weight loss from baseline of ≥5%, ≥10%, and ≥15%

- A composite endpoint "HbA1c ≤ 6.5% & Weight Loss ≥10% & No-Hypoglycemia" 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 estimand and efficacy estimand 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.

Table GPHE.4.3 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 intensified conventional care at Week 104 and at Week 208	Change from baseline in fasting serum glucose	ANCOVA model in Section 4.1.2.1 for treatment regimen estimand and MMRM model in Section 4.1.2.2 for efficacy estimand	Use baseline fasting serum glucose as a covariate.
	Both the change and percent change from baseline in lipid parameters (Total-Cholesterol, HDL-C, LDL-C, VLDL-C, TG)	ANCOVA model in Section 4.1.2.1 for treatment regimen estimand and MMRM model in Section 4.1.2.2 for efficacy estimand	Use corresponding baseline lipid parameters as a covariate.
	Proportion of patients achieving an HbA1c target value of <5.7%	Logistic model in Section 4.1.2.1 for treatment regimen and Section 4.1.2.2 for efficacy estimand	None
	Proportion of patients achieving an HbA1c target value of ≤6.5%	Logistic model in Section 4.1.2.1 for treatment regimen and Section 4.1.2.2 for efficacy estimand	None

Objective	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to Section	Additional Information
	Proportion of patients achieving an HbA1c target value of <7%	Logistic model in Section 4.1.2.1 for treatment regimen and Section 4.1.2.2 for efficacy estimand	None
	Proportion of patients who achieved weight loss of ≥5%, from baseline	Logistic model in Section 4.1.2.1 for treatment regimen and Section 4.1.2.2 for efficacy estimand	None
	Proportion of patients who achieved weight loss of ≥10%, from baseline	Logistic model in Section 4.1.2.1 for treatment regimen and Section 4.1.2.2 for efficacy estimand	None
	Proportion of patients who achieved weight loss of ≥15%, from baseline	Logistic model in Section 4.1.2.1 for treatment regimen and Section 4.1.2.2 for efficacy estimand	None
	Proportion of patients who achieved composite endpoint ("HbA1c ≤6.5% & Weight Loss ≥10% & No-Hypoglycemia"), defined as o HbA1c ≤6.5% o weight loss ≥10%, and o no hypoglycemia, defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia	Logistic model in Section 4.1.2.1 for treatment regimen and Section 4.1.2.2 for efficacy estimand	Model will include treatment and region as fixed effects, and baseline age, baseline HbA1c and baseline weight as covariates. No formal imputation method will be implemented to assess hypoglycemia.
	Change in β-cell glucose sensitivity from baseline based on the modeling analysis using Mari's model assessed in OGTT	ANCOVA model in Section 4.1.2.2	Use baseline β-cell glucose sensitivity as a covariate.

Objective	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to Section	Additional Information
	Change in insulin secretion rate at a fixed glucose concentration (adjusted and unadjusted for basal potentiation) from modeling analysis from baseline assessed in OGTT	ANCOVA model in Section 4.1.2.2	Use baseline insulin secretion rate as a covariate.
	Change in postprandial insulin sensitivity indices (Matsuda index, OGIS index) from baseline assessed in OGTT	ANCOVA model in Section 4.1.2.2	Use baseline postprandial insulin sensitivity indices (Matsuda index, OGIS index) as a covariate.
	Change in fasting and post-challenge glucose (total $AUC_{0\text{-}180min}$) from baseline assessed in OGTT	ANCOVA model in Section 4.1.2.2	Use baseline fasting and post-challenge glucose (total $AUC_{0\text{-}180min}$) as a covariate.
	Change in fasting and post-challenge glucose (incremental $AUC_{0\text{-}180min}$) from baseline assessed in OGTT	ANCOVA model in Section 4.1.2.2	Use baseline fasting and post-challenge glucose (incremental $AUC_{0\text{-}180min}$) as a covariate.
	Change in fasting and post-challenge insulin (total $AUC_{0\text{-}180min}$) from baseline assessed in OGTT	ANCOVA model in Section 4.1.2.2	Use baseline fasting and post-challenge insulin (total $AUC_{0\text{-}180min}$) as a covariate.
	Change in fasting and post-challenge insulin (incremental $AUC_{0\text{-}180min}$) from baseline assessed in OGTT	ANCOVA model in Section 4.1.2.2	Use baseline fasting and post-challenge insulin (incremental $AUC_{0\text{-}180min}$) as a covariate.
	Change in IWQOL Lite-CT Physical Functioning domain score from baseline	MMRM model in Section 4.1.2.2	Use baseline IWQOL Lite-CT PF domain score as a covariate.

Abbreviations: $AUC_{0\text{-}180min}$ = area under curve 0-180 minutes; BG = blood glucose; FSG = fasting serum glucose; HbA1c = hemoglobin A1c; HDLC = high-density lipoprotein-cholesterol; LDL-C = low-density lipoprotein cholesterol; MMRM = mixed-model repeated measure; TG = triglyceride; VLDL-C = very-low density lipoprotein-cholesterol

4.6. Exploratory Endpoints

Other exploratory (prespecified/non-prespecified in the protocol) efficacy measures 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 type 1 error rate control. Some parameters may be log transformed, if necessary.

Table GPHE.4.4 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 intensified conventional care at Week 104 and at Week 208	Change in daily average 7-point self-monitored blood glucose profiles	MMRM model in Section 4.1.2.2	Use baseline SMBG parameter as a covariate.
	Change in self-monitored blood glucose profiles • 2-hour morning • Midday • evening meal excursions • the mean of all meals 2-hour excursion • the mean of all 7-point measurements • the mean of all pre-meal measurements • the mean of all 2-hour postprandial measurements	MMRM model in Section 4.1.2.2	Use corresponding baseline parameter as a covariate.
	Change in BMI from baseline	MMRM model in Section 4.1.2.2	Use baseline BMI as a covariate.
	Both the Change and percent change in urinary albumin to creatinine ratio from baseline	MMRM model in Section 4.1.2.2	Use baseline urinary albumin to creatinine ratio as a covariate.
	Change in eGFR from baseline	MMRM model in Section 4.1.2.2	Use baseline eGFR as a covariate.
	Change from baseline in number of antihyperglycemic medications, including rescue, while patients are on treatment	Descriptive summary statistics	Additional analysis will be conducted if needed
	Time to treatment failure, defined as HbA1c ≥7% confirmed at 2 consecutive study visits after Week 32	Cox proportional hazard model using EAS	Model terms includes of treatment, region as fixed effects, and baseline HbA1c and baseline age as covariates.
	The change in glucagon concentration at fasting and post-challenge (total $AUC_{0\text{-}180min}$) from baseline assessed during OGTT	ANCOVA model in Section 4.1.2.2	Use baseline glucagon concentration at fasting and post-challenge (total $AUC_{0\text{-}180min}$) as a covariate.
	The change in glucagon concentration at fasting and post-challenge (incremental $AUC_{0\text{-}180min}$) from baseline assessed during OGTT	ANCOVA model in Section 4.1.2.2	Use baseline glucagon concentration at fasting and post-challenge (incremental $AUC_{0\text{-}180min}$) as a covariate.

Objective	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to Section	Additional Information
	The change in insulin/glucagon ratios at fasting and post-challenge (total $AUC_{0\text{-}180min}$) from baseline assessed during OGTT	ANCOVA model in Section 4.1.2.2	Use baseline insulin/glucagon ratios at fasting and post-challenge (total $AUC_{0\text{-}180min}$) as a covariate.
	The change in insulin/glucagon ratios at fasting and post-challenge (incremental $AUC_{0\text{-}180min}$) from baseline assessed during OGTT	ANCOVA model in Section 4.1.2.2	Use baseline insulin/glucagon ratios at fasting and post-challenge (incremental $AUC_{0\text{-}180min}$)e as a covariate.
	The change in C-peptide concentration at fasting and post-challenge (total $AUC_{0\text{-}180min}$) from baseline assessed during OGTT	ANCOVA model in Section 4.1.2.2	Use baseline C-peptide concentration at fasting and post-challenge (total $AUC_{0\text{-}180min}$) as a covariate.
	The change in C-peptide concentration at fasting and post-challenge (incremental $AUC_{0\text{-}180min}$) from baseline assessed during OGTT	ANCOVA model in Section 4.1.2.2	Use baseline C-peptide concentration at fasting and post-challenge (incremental $AUC_{0\text{-}180min}$) as a covariate.
	The change in insulinogenic index from baseline assessed during OGTT	ANCOVA model in Section 4.1.2.2	Use baseline insulinogenic index as a covariate.
	The change in insulin clearance from baseline assessed during OGTT	ANCOVA model in Section 4.1.2.2	Use baseline insulin clearance as a covariate.
	Change from baseline in SF-36 v2 norm-based scores: Physical component summary and Mental component summary, Physical Functioning domain score, and each additional individual domain: Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, Mental Health	MMRM model in Section 4.1.2.2	Use baseline of each corresponding norm-based score.
	Change in IWQOL-Lite-CT (Total, Physical, and Psychosocial) from baseline	MMRM model in Section 4.1.2.2	Use baseline IWQOL-Lite-CT (Total, Physical, and Psychosocial) as a covariate. I
	Change in ability to Perform Physical Activities of Daily Living (APPADL) from baseline	MMRM model in Section 4.1.2.2	Use baseline Physical Activities of Daily Living (APPADL) score as a covariate.

Objective	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to Section	Additional Information
	Change in impact of Weight on Self Perception (IW-SP)	MMRM model in Section 4.1.2.2	Use baseline impact of Weight on Self Perception (IW-SP) score as a covariate.

Abbreviations: APPADL = Ability to Perform Physical Activities of Daily Living; BMI = body mass index; FSG = fasting serum glucose; HbA1c = hemoglobin A1c; HDL-C = high-density lipoprotein-cholesterol; IW-SP = Impact of Weight on Self-Perceptions Questionnaire; LDL-C = low-density lipoprotein cholesterol; QW = once-weekly; TG = triglyceride; VLDL-C = very-low density lipoprotein-cholesterol.

Note: Exploratory Efficacy Analyses Not Controlled for Type I Error.

4.6.1. Exploratory Analysis not Specified in the Protocol

BMI-based sex-specific cardiovascular disease (CVD) risk equations developed from the Framingham Heart Study will be used to predict a patient's 10-year risk of developing a CVD event, including coronary heart disease, stroke, peripheral artery disease, or heart failure (D'Agostino et al., 2008).Complete-case analysis of participants without pre-existing CVD who complete study treatment will be conducted for the following endpoints:

- change in predicted 10-year cardiovascular risk
- percent change in predicted 10-year cardiovascular risk

The ANCOVA model will be used to analyze the defined CVD endpoints including:

- Treatment group as a factor variable
- Baseline 10-year CVD risk score as a continuous variable
- Region as a factor variable
- baseline category for HbA1c (≤8.0%,>8.0%), and
- Age as a continuous variable.

The marginal Hazard Ratio (HR) for treatment effect will be estimated as the ratio of the negative logarithms of the group-level predicted 10-year survival probabilities, under the proportional hazards assumption.

Subgroup analyses will be conducted for the defined CVD endpoints. The ANCOVA model specified above will be fitted separately within each subgroup below:

- sex: individual AFAB, individual AMAB, and
- baseline BMI: <35, ≥35 kg/m^2.

Additional analysis of non-protocol specified exploratory efficacy endpoints are given below.

Table GPHE.4.5 Non- protocol 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 intensified conventional care at Week 104 & Week 208	• Both the change and percent change from baseline in the following parameters assessed during OGTT: ○ Insulin: HOMA-S, HOMA-IR, HOMA-B	ANCOVA model in Section 4.1.2.2	Use corresponding baseline insulin values as a covariate.
	• Both the change and percent change from baseline in the following parameters assessed during OGTT: ○ C-peptide: HOMA-S, HOMA-IR, HOMA-B	ANCOVA model in Section 4.1.2.2	Use corresponding baseline C-peptide values as a covariate.
	• Percentage of participants who achieved the following composite endpoints where Hypoglycemia is defined as BG <54 mg/dL (<3.0 mmol/L) and/or severe hypoglycemia: ○ HbA1c <5.7% & Weight Loss ≥15% & No-Hypoglycemia ○ HbA1c ≤6.5% & Weight Loss ≥10% & No-Hypoglycemia	Logistic model in Section 4.1.2.2	None
	• Number of antihypertensive medications	Descriptive summary statistics	Additional analysis will be conducted if needed
	• Number of lipid-lowering medications	Descriptive summary statistics	Additional analysis will be conducted if needed
	• Change from baseline in high sensitivity (hsCRP) and microalbumin	MMRM model in Section 4.1.2.2	Use corresponding baseline parameter as a covariate.

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 intensified conventional care at Week 208	• HbA1c <5.7%, ≤6.5%, and <7% in participants achieving HbA1c <5.7%, ≤6.5%, and <7% respectively, at Week 104	Logistic model in Section 4.1.2.2	None
	• Weight loss of ≥5%, ≥10%, and ≥15% in participants achieving Weight loss of ≥5%, ≥10%, and ≥15% respectively, at Week 104	Logistic model in Section 4.1.2.2	None
	• HbA1c ≤6.5% & Weight Loss ≥10% & No-Hypoglycemia in patients achieving HbA1c ≤6.5% & Weight Loss ≥10% & No-Hypoglycemia at Week 104	Logistic model in Section 4.1.2.2	None
	• HbA1c ≤5.7% & Weight Loss ≥15% & No-Hypoglycemia in patients achieving HbA1c ≤5.7% & Weight Loss ≥15% & No-Hypoglycemia at Week 104	Logistic model in Section 4.1.2.2	None

4.7. Safety Analyses

Unless specified otherwise, safety assessments will be conducted based on the Safety Analysis Set (see Section 3) irrespective of adherence to study intervention or initiation of rescue antihyperglycemic therapy. Selected safety analyses may be conducted after excluding data on rescue therapy or data after starting another antihyperglycemic medication.

4.7.1. Extent of Exposure

Duration of exposure to study intervention will be summarized by treatment group for safety participants. Exposure will be summarized and calculated for the treatment period being considered as the date of last dose of study intervention minus the date of first dose of study intervention plus 1 day using safety participants. Descriptive statistics (including n, mean, SD, minimum, first quartile, median, third quartile, and maximum) will be provided. No p-values will be reported in these summaries as they are intended to describe the study populations rather than test hypotheses about populations.

4.7.2. Adverse Events

Adverse events will be coded from the actual term using the Medical Dictionary for Regulatory Activities and reported with preferred terms and system organ class.

A TEAE is defined as medical occurrence that emerges during a defined treatment period, having been absent pre-treatment, or worsens relative to the pre-treatment 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. For events with a missing severity during the baseline period, it 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 preferred term (PT) nested within system organ class (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, serious adverse event (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 non-missing 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.

4.7.2.1. Deaths

A listing of all deaths will be provided. The listing will include patient identification 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 (CEC).

4.7.2.2. 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

Approved on 19 Nov 2025 GMT

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 identification 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.

4.7.2.3. 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 MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC. A listing of all the discontinuation from Study due to Adverse Event will be provided.

4.7.2.4. 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 Adverse Event will be provided.

4.7.3. Special Safety Topics

4.7.3.1. 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 blood glucose-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 blood glucose values ≤70 mg/dL (3.9 mmol/L) occurring within a 1-hour period may be considered to be a single hypoglycemic event and the one with the lowest blood glucose value can be selected to be the representative. If two records with the same lowest values within an 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, level 1 and level 2 hypoglycemia, and clinically significant documented symptomatic level 1 and level 2 hypoglycemia (defined as level 1 or 2 hypoglycemia with symptoms and/or signs of hypoglycemia, respectively), 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 (tirzepatide 15 mg or MTD or the intensified conventional care), baseline HbA1C category (≤8.0%,>8.0%), and region as fixed effects and Age as a covariate. The rate of hypoglycemic episodes per patient year may be analyzed using a generalized linear mixed-effects model assuming the number of hypoglycemic episodes follows a negative binomial distribution with the mean modeled using region, baseline HbA1C category (≤8.0%, >8.0%), and treatment as fixed effects and Age as a covariate 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.

4.7.3.2. Severe, Persistent Hyperglycemia

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

4.7.3.3. 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 pre-defined SMQ search for acute pancreatitis and MedDRA PT of pancreatitis chronic. Detailed searching criteria can be found in Appendix 5 (Section 6.5).

4.7.3.3.1. Pancreatic Hyperenzymeia

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.

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

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

4.7.3.4.1. 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.

4.7.3.5. 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.

4.7.3.6. Arrhythmias and Cardiac Conduction Disorders

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

4.7.3.7. 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 is collected, the events occurred on the same date as the study drug injection date will be included.
- **Potential Non-Immediate 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 Appendix 5 (Section 6.5).

4.7.3.8. 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 pre-defined 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 Appendix 5 (Section 6.5). The PT will be used for summary by treatment within each HLT category.

4.7.3.9. Diabetic Retinopathy Complications

Results of the baseline dilated fundoscopic exam will be summarized by treatment and by visit at Baseline, Week 104 and Week 208. 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. 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/serious adverse events from the PTs defined in searching criteria in Appendix 5 (Section 6.5) will be summarized.

4.7.3.10. Hepatic Safety

4.7.3.10.1. Hepatobiliary Disorders

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

4.7.3.10.2. Acute Gallbladder Disease

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

4.7.3.10.3. Liver Enzymes

Analyses for laboratory analyte measurements are described in Section 4.7.5. 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 to <3 × ULN, ≥3 to <5 × ULN, ≥5 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 to <3 × ULN, ≥3 to <5 × ULN, ≥5 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 (≤1 × ULN, >1 to <1.5 × ULN, ≥1.5 to <2 × ULN, ≥2 ULN) to postbaseline with the following categories: ≤1 × ULN, > 1 to <2 × ULN, ≥2 × ULN.
- Shift tables of serum alkaline phosphatase from baseline (≤1 × ULN, >1 to <1.5 × ULN, ≥1.5 to <2 × ULN, ≥2 ULN) to postbaseline with the following categories: ≤1 × ULN, >1 to <2 × ULN, ≥2 × ULN.

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

4.7.3.11. 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.

4.7.4. 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 geographical region, baseline HbA1c (≤8.0%, >8.0%), treatment group, visit, and treatment-by-visit interaction 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 GPHE.4.6.

Table GPHE.4.6 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

Abbreviations: BP = blood pressure; bpm = beats per minute.

4.7.5. 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.

4.7.6. 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 SAP had intervention not been taken will be provided. Additional summaries will be provided as deemed appropriate.

4.8. Other Analyses

4.8.1. Subgroup Analyses

Subgroup analyses of the primary endpoint, change from baseline in HbA1c, at both W104 and W208, will be made based on the efficacy estimand using the EAS to assess consistency of the intervention effect across the following subgroups:

- Age group: ≤ 40 years vs ≥ 41 years
- Sex: Female vs Male
- Race: White vs Black vs Other
- Ethnicity: Hispanic vs not Hispanic
- Duration of diabetes (≤2 years, >2 year)
- Baseline BMI (<27, ≥27, <30, ≥30, and <35, ≥35 kg/m^2),
- eGFR (<60, ≥60)
- Baseline HbA1c (≤8.0%, >8.0%), and
- Region: NA, LATAM, EU+Israel.

4.8.1.1. Subgroup Analysis of Weight Change at Week 104 and Week 208

Subgroup analyses by the following baseline characteristics will be provided for weight change from baseline at Week 104 and Week 20 based on the efficacy estimand using EAS:

- Age group (≤40, ≥41 years)
- Sex: female vs male
- Baseline BMI (<27, ≥27, <30, ≥30, and <35, ≥35 kg/m2), and
- Baseline HbA1c (≤8.0%, >8.0%).

4.9. Interim Analyses

An interim analysis for primary outcome will be performed when all participants complete 104 weeks of treatment or discontinue. The final analysis will be performed when all participants complete the study.

Approved on 19 Nov 2025 GMT

4.10. Changes to Protocol-Planned Analyses

There are no major changes to the protocol-planned analyses that are proposed in this statistical analysis plan. Most of them are related to minor clarifications related to imputation strategy and editorial changes.

5. Sample Size Determination

Although the primary objective of the trial is to demonstrate that tirzepatide 15 mg or MTD is non-inferior to intensified conventional care for the primary endpoint (change in HbA1c from baseline to Week 104) with a 0.3% non-inferiority margin, sample size selection is guided by establishing superiority of tirzepatide 15 mg or MTD compared to intensified conventional care for change in HbA1c from baseline to Week 104 and to Week 208 (key secondary objectives), irrespective of adherence to treatment or introduction of antihyperglycemic rescue medication (treatment-regimen estimand).

A sample size of 780 participants (390 per group) will provide at least 90% power to detect a difference between treatment groups at both Week 104 and Week 208 irrespectively of adherence to treatment, changes in background metformin medication, or introduction of antihyperglycemic rescue medication.

This sample size is based on a treatment difference of 0.33% at Week 104 and 0.38% at Week 208 with a common standard deviation of 1.3% and 1.4%, respectively, using a two-group t-test and a 0.05 two-sided significance level. The treatment differences are based on the following assumptions:

- Efficacy observed for intensified conventional care at Weeks 104 and 208 will be similar to that of semaglutide 1.0 (50% of the participants) and semaglutide 2.0 (50% of the participants)
- Tirzepatide 15 mg or MTD will be comprised of these tirzepatide doses
    - 15 mg dose: 80% of the participants
    - 10 mg dose: 15% of the participants, and
    - 5 mg dose: 5% of the participants
- Increases of HbA1c over time
    - intensified conventional care will show a 0.25% increase in HbA1c at Week 104 and 0.4% at Week 208, and
    - Tirzepatide will show a 0.2% increase in HbA1c at Week 104 and 0.3% at Week 208.

In addition, the planned sample size provides at least 90% power to detect a difference between treatment groups at both Week 104 and Week 208 after randomization regardless of changes in background metformin medication, assuming that participants had stayed on treatment and not taken antihyperglycemic rescue medication.

This sample size is based on a treatment difference of 0.33% at Week 104 and 0.38% at Week 208 with a common standard deviation of 1.1%, using a two-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 the following discontinuation rates

- 25% of the participants discontinue study intervention by Week 104, and
- 35% of the participants discontinue study intervention by Week 208.

Approved on 19 Nov 2025 GMT

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 randomized 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 T2DM, and estimated glomerular filtration rate (eGFR).

6.2. Appendix 2: Treatment Compliance

Overall treatment compliance will be defined as taking at least 75% of the scheduled tirzepatide/intensified conventional care doses. Compliance for the intensified conventional care arm will be calculated by calculating the compliance of each individual study drug the participant is prescribed during the study and then averaging the individual compliance rates. Compliance will be calculated by taking the number of doses administered (regardless of the actual dose administered) divided by the total number of doses expected to be administered ×100. Treatment compliance will be summarized descriptively 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:

- baseline metformin total daily dose
- 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, and
    - antiemetic medication.

6.4. Appendix 4: Clinical Trial Registry

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:

Approved on 19 Nov 2025 GMT

- Summary of adverse events, provided as a dataset which will be converted to an XML file. Both Serious Adverse Events and 'Other' Non-Serious Adverse Events are summarized: by treatment group, by MedDRA preferred term.
    - An adverse event is considered 'Serious' whether or not it is a treatment emergent adverse event (TEAE).
    - An adverse event is considered in the 'Other' category if it is both a TEAE and is not serious. For each Serious AE 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 Serious AE, 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 (0-27 days), infants and toddlers (28 days – 23 months, children (2-11 years), adolescents (12-17 years), adults (18-64 years), 65-85 years, and 85 years and over.

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

The adverse events of special interest (AESI) analyses are detailed in Section 4.7.3. The search criteria for each AESI are stored in CLUWE: \\statsclstr.am.lilly.com\lillyce\prd\ly3298176\common\AESI_Lab\Search criteria AESIs_TZP.xlsx.

Approved on 19 Nov 2025 GMT

7. References

Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med*. 2014;33(4):693-713. https://doi.org/10.1002/sim.5974

Bretz F, Maurer W, Brannath W, 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

Bretz F, Posch M, Glimm E, et al. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J*. 2011;53(6):894-913. https://doi.org/10.1002/bimj.201000239

Dahl D, Onishi Y, Norwood P, et al. Effect of subcutaneous tirzepatide vs placebo added to titrated insulin glargine on glycemic control in patients with type 2 diabetes: The SURPASS-5 randomized clinical trial. *JAMA*. 2022;327(6):534–545. https://doi.org/10.1001/jama.2022.0078

Del Prato S, Kahn SE, Pavo I, et al. Tirzepatide versus insulin glargine in type 2 diabetes and increased cardiovascular risk (SURPASS-4): a randomised, open-label, parallel-group, multicentre, phase 3 trial. *Lancet*. 2021;398(10313):1811-1824. https://doi.org/10.1016/S0140-6736(21)02188-7

Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford: Clarendon Press; 1994.

[FDA] United States Food and Drug Administration. E9(R1) Statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials: guidance for industry. Published May 2021. Accessed September 12, 2023. https://www.fda.gov/media/148473/download

Frías JP, Davies MJ, Rosenstock J, et al. Tirzepatide versus semaglutide once weekly in patients with type 2 diabetes. *N Engl J Med*. 2021;385(6):503-515. https://doi.org/10.1056/NEJMoa2107519

Ludvik B, Giorgino F, Jódar E, et al. Once-weekly tirzepatide versus once-daily insulin degludec as add-on to metformin with or without SGLT2 inhibitors in patients with type 2 diabetes (SURPASS-3): a randomised, open-label, parallel-group, phase 3 trial. *Lancet*. 2021;398(10300):583-598. https://doi.org/10.1016/S0140-6736(21)01443-4

Mari A, Tura A, Gastaldelli A, Ferrannini E: Assessing insulin secretion by modeling in multiple-meal tests: role of potentiation. *Diabetes*. 2002;51(Suppl 1):S221-S226, 2002. https://doi.org/10.2337/diabetes.51.2007.s221

Rosenstock J, Wysham C, Frías JP, et al. Efficacy and safety of a novel dual GIP and GLP-1 receptor agonist tirzepatide in patients with type 2 diabetes (SURPASS-1): a double-blind, randomised, phase 3 trial. Lancet. 2021;398(10295):143-155. https://doi.org/10.1016/S0140-6736(21)01324-6.

Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons Inc.; 1987.

Wang B, Du Y. Robustly leveraging post-randomization information to improve precision in randomized trials. *Mathematics*. 2022. https://arxiv.org/abs/2110.09645v2

Signature Page for VV-CLIN-056527 v2.0

Approval	PPD Statistician 19-Nov-2025 20:03:11 GMT+0000

Signature Page for VV-CLIN-056527 v2.0