Statistical Analysis Plan J2A-MC-GZGQ Version 3.0

A Phase 3, Randomized, Double-Blind Study to Investigate the Efficacy and Safety of Once-Daily Oral LY3502970 Compared With Placebo in Adult Participants With Obesity or Overweight and Type 2 Diabetes (ATTAIN-2)

NCT05872620

Approval Date: 26-June-2025

Title Page

Protocol Title: A Phase 3, Randomized, Double-Blind Study to Investigate the Efficacy and Safety of Once-Daily Oral LY3502970 Compared with Placebo in Adult Participants with Obesity or Overweight and Type 2 Diabetes (ATTAIN-2)

Protocol Number: J2A-MC-GZGQ

Compound Number: LY3502970 (orforglipron)

Short Title: GZGQ SAP

Acronym: ATTAIN-2

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana USA 46285

Regulatory Agency Identifier Numbers:

Registry	ID
IND:	156143
EU trial number	2022-502837-24-00

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-117489

Approved on 26 Jun 2025 GMT

Table of Contents

Table of Contents 2
Version History 6
1. Introduction 11
1.1. Objectives, Endpoints, and Estimands 12
1.2. Study Design 17
2. Statistical Hypotheses 20
2.1. Multiplicity Adjustment 21
3. Analysis Sets 23
4. Statistical Analyses 25
4.1. General Considerations 25
4.1.1. Definition of Baseline 26
4.1.2. Analysis Methods 26
4.2. Participant Dispositions 34
4.3. Primary Endpoint/Estimand Analysis 35
4.3.1. Definition of Endpoint(s) 35
4.3.2. Sensitivity Analyses 35
4.3.3. Main Analytical Approach 35
4.3.4. Supplementary Analyses 35
4.4. Secondary Endpoints/Estimands Analysis 35
4.4.1. Key Secondary Endpoints 52
4.4.2. Additional Secondary Endpoints 52
4.5. Exploratory Endpoints/Estimands Analysis 52
4.6. Safety Analyses 52
4.6.1. Extent of Exposure 53
4.6.2. Adverse Events 53
4.6.3. Clinical Laboratory Evaluation 54
4.6.4. Vital Signs and Physical Characteristics 54
4.6.5. Electrocardiograms 54
4.6.6. Safety Topics of Interest 55
4.7. Other Analyses 63
4.7.1. Subgroup Analyses 63
4.8. Interim Analyses 64
4.8.1. Data Monitoring Committee 64
4.9. Changes to Protocol-Planned Analyses 64
5. Sample Size Determination 65
6. Supporting Documentation 66

Approved on 26 Jun 2025 GMT

6.1. Appendix 1: Demographic and Baseline Characteristics 66
6.2. Appendix 2: Historical Illnesses and Preexisting Conditions 67
6.3. Appendix 3: Prior/Concomitant Medications....68
6.4. Appendix 4: Treatment Compliance 70
6.5. Appendix 5: Important Protocol Deviations....71
6.6. Appendix 6: Clinical Trial Registry Analyses....71
6.7. Appendix 7: Statistical Analysis for China 71
6.8. Appendix 8: Statistical Analysis for Japan....72
7. References 74

Approved on 26 Jun 2025 GMT

Table of Contents

Tables **Page**

Table GZGQ.3.1. Participant Analysis Sets....23
Table GZGQ.3.2. Data Points Sets for Non–Glycemic-Related Parameters....23
Table GZGQ.3.3. Data Points Sets for Glycemic-Related Parameters....24
Table GZGQ.3.4. Data Points Sets for Safety Analyses....24
Table GZGQ.4.1. Baseline and Postbaseline Definitions for Safety Analyses....26
Table GZGQ.4.2. Imputation Procedure....29
Table GZGQ.4.3. Categorization of Treatment Period Discontinuation Reasons....30
Table GZGQ.4.4. Listings and Summary Tables Related to Dispositions....34
Table GZGQ.4.5. Description and Derivation of Efficacy Measures and Patient Reported Outcomes....36
Table GZGQ.4.6. Description of Efficacy/Health Outcomes Analyses....47
Table GZGQ.4.7. Summary Tables Related to Adverse Events....53
Table GZGQ.4.8. Summary Tables Related to Clinical Laboratory Evaluations....54
Table GZGQ.4.9. Summary Tables Related to Vital Signs....54
Table GZGQ.4.10. Summary Tables Related to ECG Parameters....55
Table GZGQ.4.11. Description and Analyses of Safety Topics of Interest....56
Table GZGQ.6.1. Demographics and Baseline Characteristics with Variables for Subgroup Analysis....66
Table GZGQ.6.2. Summary Tables Related to Historical Illness and Preexisting Conditions....68
Table GZGQ.6.3. Summary Tables Related to Concomitant Medications....68
Table GZGQ.6.4. Protocol-Specified Concomitant Medications....69
Table GZGQ.6.5. Prohibited Concomitant Medication Rules....69
Table GZGQ.6.6 The 11 Obesity-Related Health Disorders....73

Approved on 26 Jun 2025 GMT

Table of Contents

Figure **Page**

Figure GZGQ.1.1. Illustration of study design for clinical protocol J2A-MC-GZGQ. 19

Figure GZGQ.2.1. Graphical testing procedure. 22

Figure GZGQ.4.1. Treatment journey of a trial participant. 28

Version History

This statistical analysis plan (SAP) for study J2A-MC-GZGQ (GZGQ) is based on the protocol amendment (d) dated 31 October 2023.

The GZGQ SAP Version 2 included updates based on regulatory feedback and protocol. The major updates in the Version 2 amendment are listed in the table below.

GZGQ SAP Version 3 was updated based on regulatory feedback. The major updates in this amendment are listed in the table below.

SAP Version History Summary

SAP Version	Approval Date	Section	Change	Rationale
1	January 24, 2024		Not Applicable	Original version
2	November 6, 2024	3	Removed note regarding no inferential comparisons for the summary of disposition, demographics, historical illness, and preexisting conditions	Clarification
		4.1.2.1.2	Further clarified the definition of Scenario 5. Removed the direct imputation method Updated the imputation method for Scenario 5B Added footnote in Table GZGQ.4.4 for protocol deviation, inadvertent enrolment, and emergency unblinding	The direct imputation method is not fully aligned with the primary imputation strategy. Used a different method to handle Scenario 5B, based on simulation results from Liu et al. (2024) Clarified the distinction of imputing discontinuation due to protocol deviation, inadvertent enrolment

		4.1.2.1.4	Updated the penalty range for the MI-TP analysis to -15% to 15%	Updated due to magnitude of projected weight loss
		4.1.2.3	Updated MMRM model for safety analyses Clarified the definition of start time for different analyses	Simplify the analysis model for safety analyses To be consistent with the PSAP
		4.6.6	Modified Table GZGQ.4.11	To be consistent with the PSAP
		4.7.1	Updated the testing of subgroup effects. Added details for the Bayesian shrinkage method	Clarification
		Table GZGQ.6.1	Updated subgroup analyses	Clarification
		6.7	Updated the statistical analysis for China	Clarification
		7	Reference section was updated	To add extra references
3	See date on Page 1	1.1	Updated language on PGI endpoints	Clarification
		1.1	Updated language on Population definition	Clarification
		1.1	Updated language on the ICE definitions for glycemic parameters	Clarification

Approved on 26 Jun 2025 GMT

		1.1 2	Moved 6 mg and 12 mg endpoint "mean change in waist circumference (cm)" from key secondary list to additional secondary list, and updated hypotheses list accordingly.	Limited clinical meaningfulness in changes in waist circumference from 6 mg and 12 mg.
		2.1	Updated the graphical approach scheme with the final version	Clarification
		Table GZGQ 3.4	Updated the data points sets for safety analyses	Clarification
		4.1	Replaced country with region in the statistical analysis models	To improve model convergence
		Table GZGQ 4.1	Minor word changes in Baseline and Postbaseline definition for safety analyses	Clarification
		4.1.2.1.1	For method for continuous variables, added description for obtaining estimates of the pooled dose vs. placebo	Clarification
		4.1.2.1.2	Clarified the imputation strategy for all scenarios Updated study discontinuation reason categories for imputation purpose	Clarification
		4.1.2.1.3	Minor word changes in modified multiple imputation strategy	Clarification

Approved on 26 Jun 2025 GMT

		4.1.2.1.5	For the method for binary variables, added interaction terms in logistic regression, and clarified the statistical testing for risk difference is the primary approach for treatment comparisons	Consistency with other analyses Clarification
		4.1.2.2.1	Clarified that maximum likelihood based MMRM is used for continuous variables for efficacy estimand	Clarification
		4.1.2.3	Added that for smaller number of events negative binomial will be replaced by empirical method for analysis of rate of hypoglycemia	The approach aligns if there are limited events.
		4.4	In Table GZGQ.4.5, modified and added variables	Clarification and to align with study endpoints
		4.4	In Table GZGQ.4.6, modified and added additional analyses	Clarification and to align with study objectives
		4.6.2	In Table GZGQ.4.7, modified analyses	Clarification and to align with study objectives
		4.6.6	In Table GZGQ. 4.11, modified and added additional analyses	To be consistent with PSAP
		4.7.1	Added details for the Bayesian shrinkage	Clarification

			method for subgroup analysis	
		6.3	In Table GZGQ.6.3, modified and added additional analyses	Clarification and to align with study objectives
		6.7, Appendix 7	Modified parameter list and provide additional details on analyses method	Clarification
		Throughout SAP	Minor changes and reorganization	For clarity, no change to analysis methodologies, so not detailed

Abbreviations: ICE = intercurrent event; PGI = Patient Global Impression; PSAP= program safety analysis plan; SAP = statistical analysis plan, MI-TP = multiple imputation based tipping-point analysis; vs = versus.

1. Introduction

This statistical analysis plan (SAP) is intended to describe the analyses of primary and secondary objectives, as well as safety assessments for Study J2A-MC-GZGQ (GZGQ). Sensitivity and supplementary analyses intended to support the primary and key secondary objectives are also included. Additional exploratory analyses, if needed, may be included in a separate document.

Pharmacokinetic/pharmacodynamic analyses will be documented in a separate document.

Changes to the protocol-planned analyses, if any, are described in Section 4.9.

Approved on 26 Jun 2025 GMT

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Primary	
To demonstrate that orforglipron 6 mg, 12 mg, and/or 36 mg is superior to placebo in change from baseline for the following: • Body weight	From baseline to Week 72, • Mean percent change in body weight
Key Secondary (Controlled for Type 1 Error)	
To demonstrate that orforglipron 6 mg is superior to placebo in change from baseline for the following: • Body weight	From baseline to Week 72, • Percentage of participants who achieve a body weight reduction of o ≥5% o ≥10%
To demonstrate that orforglipron 12 mg and/or 36 mg is superior to placebo in change from baseline for the following: • Body weight	From baseline to Week 72, • Percentage of participants who achieve a body weight reduction of o ≥5% o ≥10% o ≥15%
To demonstrate that orforglipron 36 mg is superior to placebo in change from baseline for the following: • Waist circumference	From baseline to Week 72, • Mean change in waist circumference (cm)
To demonstrate that orforglipron 6 mg, 12 mg, and/or 36 mg is superior to placebo in change from baseline for the following: • Glycemic control	From baseline to Week 72, • Mean change in HbA1c (%) • Mean change in fasting glucose (mg/dL) • Percentage of participants achieving HbA1c target value of <7.0% • Percentage of participants achieving HbA1c target value of ≤6.5%
To demonstrate that orforglipron (6-mg, 12-mg, and 36-mg pooled doses) is superior to placebo in change from baseline for the following: • SBP • Lipid parameters	From baseline to Week 72, • Mean change in SBP (mmHg) • Mean percent change in fasting o non-HDL cholesterol o triglycerides

Approved on 26 Jun 2025 GMT

Objectives	Endpoints
Additional Secondary	
To demonstrate that orforglipron 6 mg, 12 mg, and/or 36 mg is superior to placebo in change from baseline for the following: • Body weight • Glycemic control • Fasting insulin	From baseline to Week 72, • Mean change in absolute body weight (kg) • Mean change in BMI (kg/m^2) • Percentage of participants achieving HbA1c target value of <5.7% • Mean percent change in fasting insulin
To demonstrate that orforglipron 6 mg and/or 12 mg is superior to placebo in change from baseline for the following: • Waist circumference	From baseline to Week 72, • Mean change in waist circumference (cm)

To demonstrate that orforglipron (6-mg, 12-mg, and 36-mg pooled doses) is superior to placebo in change from baseline for the following:
• DBP
• Lipid parameters
• Patient-reported outcomes
	From baseline to Week 72,
	• Mean change in DBP (mmHg)
	• Mean percent change in fasting
	  ○ total cholesterol
	  ○ LDL-cholesterol
	  ○ HDL cholesterol
	From baseline to Week 72,
	• Mean change in SF-36v2 acute form domain scores
	• Mean change in EQ-5D-5L health state utilities and CCI
	• Mean change in IWQOL-Lite-CT Physical Function, Physical, and Psychosocial composite scores, and total score
	• Proportion of participants with improved categorical shift in PGI-S physical function due to weight
	• Proportion of participants with improvement in PGI-C physical function due to weight

To describe the safety of orforglipron as compared to placebo	Frequency of treatment-emergent adverse events
Tertiary	
To characterize the population PK of orforglipron and explore the relationships between orforglipron concentration and efficacy, safety, and tolerability measures	Population PK and PD parameters

Abbreviations: BMI = body mass index; HbA1c = hemoglobin A1c; DBP = diastolic blood pressure; EQ-5D-5L = European Quality of Life-5 dimensions-5 levels; HDL = high-density lipoprotein; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite Clinical Trials; LDL = low-density lipoprotein; PD = pharmacodynamics; PGI-C = Patient Global Impression of Change; PGI-S = Patient Global Impression of

Severity; PK = pharmacokinetics; SBP = systolic blood pressure; SF-36v2 = Short Form 36 Version 2 health survey acute form; CCI

Estimands

There will be 2 estimands for the primary objective planned in the study. The estimands address intercurrent events (ICEs) using either the treatment policy strategy or the hypothetical strategy (ICH 2019), respectively.

Estimands for weight-related, blood pressure, and lipid parameters

The below estimands will be applied to weight-related, blood pressure (BP) and lipid parameters, using percent change in body weight at Week 72 visit as an example. For other weight-related, BP, and lipid parameters, replace percent change in body weight by the endpoints to be analyzed.

Treatment regimen estimand

The treatment regimen estimand will be the primary estimand.

The clinical question of interest: *What is the treatment difference in the percent change in body weight from baseline to 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg vs. placebo, as an adjunct to healthy diet and physical activity, in individuals who meet eligibility criteria regardless of adherence to study intervention or initiation of prohibited weight management treatments?*

Treatment policy strategy

The occurrence of the ICE is considered irrelevant in defining the treatment effect of interest; the values for the variable of interest are used regardless of whether the ICE occurs.

The "treatment regimen" estimand is described by the following attributes:

- ***Population:*** Individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
- ***Endpoints:*** Percent change from baseline in body weight at 72 weeks.
- ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications, regardless of adherence to study intervention or initiation of prohibited weight management treatments.
- ***ICEs:*** No ICEs are defined because treatment adherence and the initiation of prohibited weight management treatments are a part of the treatment condition.
- ***Population-level summary and treatment effect of interest:*** Difference in mean percent changes from baseline in body weight at 72 weeks between orforglipron and placebo.

Rationale for the estimand: The estimand aims to evaluate the efficacy of orforglipron that reflects the real-life behavior of the target population.

Efficacy estimand

The clinical question of interest: *What is the treatment difference in the percent change in body weight from baseline to 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg vs. placebo, as an adjunct to healthy diet and physical activity, in individuals who meet the eligibility criteria if they would remain on their randomly assigned treatment for 72 weeks and would not initiate prohibited weight management treatments?*

Hypothetical strategy

A scenario is envisaged in which the ICE would not occur. The value of the variable to reflect the clinical question of interest is the value that the variable would have taken in the hypothetical scenario defined.

The "efficacy" estimand is described by the following attributes:

- ***Population:*** Individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
- ***Endpoints:*** Percent change in body weight from baseline at 72 weeks.
- ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications.
- ***ICEs:*** ICEs include permanent discontinuation of study intervention and initiation of prohibited weight management treatments, which is handled by the hypothetical strategy. The potential outcome of interest is the response in the efficacy measurement if participants would remain on their randomly assigned study intervention for 72 weeks and would not initiate prohibited weight management treatments. Dose modification and interruption will not be considered an ICE because they are part of the treatment condition.
- ***Population-level summary and treatment effect of interest:*** Difference in mean percent changes in body weight from baseline at 72 weeks between orforglipron and placebo.

Rationale for the estimand: This estimand aims to evaluate the efficacy of orforglipron under the ideal condition that all participants would adhere to the randomly assigned study intervention without being confounded by the initiation of prohibited weight management treatments.

Estimands for glycemic parameters

The below estimands will be applied to glycemic parameters, using change in hemoglobin A1c (HbA1c) at Week 72 as an example. For other glycemic parameters, replace change in HbA1c by the endpoints to be analyzed.

Treatment regimen estimand

The treatment regimen estimand will be the primary estimand.

The clinical question of interest: *What is the treatment difference in the change in HbA1c from baseline at 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg vs. placebo, as an adjunct to*

healthy diet and physical activity, in individuals who meet the eligibility criteria regardless of adherence to study intervention or initiation of prohibited weight management treatments or glycemic rescue therapy?

The "treatment regimen" estimand is described by the following attributes:

- ***Population:*** Individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
- ***Endpoints:*** Change in HbA1c from baseline to 72 weeks.
- ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications regardless of adherence to study intervention or initiation of prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic therapy.
- ***ICEs:*** No ICEs are defined because treatment adherence and the initiation of prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic therapy are a part of the treatment condition.
- ***Population-level summary and treatment effect of interest:*** Difference in mean changes in HbA1c from baseline to Week 72 between orforglipron and placebo.

Rationale for the estimand: The estimand aims to evaluate the efficacy of orforglipron that reflects the real-life behavior of the target population.

Efficacy estimand

The clinical question of interest: *What is the treatment difference in the change in HbA1c from baseline at 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg vs. placebo, as an adjunct to healthy diet and physical activity, in individuals who meet the eligibility criteria if they would remain on their randomly assigned treatment for 72 weeks and would not initiate prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic therapy?*

The "efficacy" estimand is described by the following attributes:

- ***Population:*** Individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
- ***Endpoint:*** Change in HbA1c from baseline to 72 weeks.
- ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications.
- ***ICEs:*** ICEs include permanent discontinuation of study intervention and initiation of prohibited weight management treatments, or glycemic rescue therapy or prohibited glycemic therapy, which is handled by the hypothetical strategy. The potential outcome of interest is the response in the efficacy measurement if participants would remain on their randomly assigned study intervention for 72 weeks and would not initiate prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic

Approved on 26 Jun 2025 GMT

therapy. Dose modification and interruption will not be considered as ICEs because they are a part of the treatment condition.

- ***Population-level summary and treatment effect of interest:*** Difference in mean changes in HbA1c from baseline to Week 72 between orforglipron and placebo.

Rationale for the estimand: This estimand aims to evaluate the efficacy of orforglipron under the ideal condition that all participants would adhere to the randomly assigned study intervention without being confounded by the initiation of prohibited weight management treatments or glycemic rescue therapy or prohibited glycemic therapy.

1.2. Study Design

Study GZGQ is a Phase 3, multicenter, randomized, parallel-arm, double-blind, placebo-controlled, 72-week study. The study will investigate the safety and efficacy of treatment with daily oral doses of orforglipron (6 mg, 12 mg, and 36 mg) compared with placebo, in participants with obesity (body mass index [BMI] 30 kg/m^2 or greater) or overweight (BMI 27 kg/m^2 or greater) who have type 2 diabetes (T2D). Unblinding details are specified in the blinding and unblinding plan.

Participants may be treated with up to 3 oral AHMs (excluding dipeptidyl peptidase 4 inhibitor [DPP-4i] and glucagon-like peptide-1 receptor agonist [GLP-1 RA]) or may be treated with diet and exercise alone; regardless of type of treatment (pharmacotherapy or diet and exercise), it must be stable for at least 90 days prior to Visit 1.

Study participants will be randomly assigned in a 1:1:1:2 ratio to receive daily dose of orforglipron (6 mg, 12 mg, or 36 mg) or placebo. An upper limit of 70% enrollment of females will be used to ensure a sufficiently large sample of males. In addition, an upper limit of 30% enrollment of participants treated with sulfonylurea will be used to allow sufficient enrollment of participants treated with other oral AHMs.

- Screening period: 3 weeks
- Treatment period
  - Dose escalation period: up to 20 weeks
  - Maintenance dose period: at least 52 weeks
- Post-treatment follow-up period: 2 weeks

Study intervention dose reduction during the entire course of the study is only permitted for management of intolerable gastrointestinal (GI) symptoms (after implementation of symptomatic treatment and of a temporary drug interruption). If study intervention is reduced, then only 1 re-escalation attempt is permitted during the entire course of the study, and only 1 de-escalation is permitted after re-escalation. If intolerable GI symptoms persist after the final dose de-escalation, the participant should be permanently discontinued from the study intervention.

The study duration will be approximately 77 weeks, including screening and follow-up. The treatment duration will be approximately 72 weeks. The visit frequency will be every 4 weeks. The study schema is shown in Figure GZGQ.1.1.

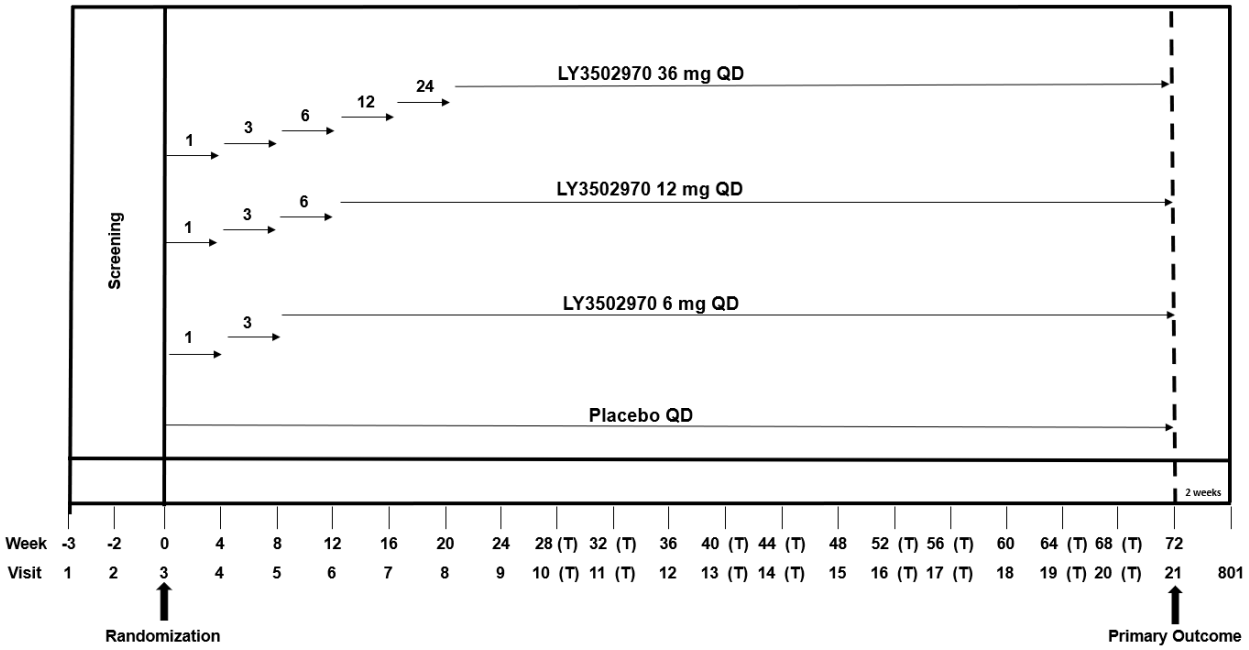
The randomization will be stratified by

- country
- sex
    - female
    - male
- background oral AHMs classified according to potential effect on body weight, see table below.

Stratification based on the use of antihyperglycemic medications

Stratum	Potential Body Weight Effect	Participants to be Included in the Corresponding Stratum
1	Significant weight loss	• The use of a SGLT-2 inhibitor alone or in combination with any other AHMs, except for an SU or TZD
2	Significant weight gain	• The use of an SU alone or in combination with any other AHMs, except for an SGLT-2 inhibitor OR • The use of a TZD alone or in combination with any other AHMs, except for an SGLT-2 inhibitor
3	Weight neutral and others	• Lifestyle management only: no background AHMs OR • Metformin monotherapy OR • The use of an SGLT-2 inhibitor in combination with an SU and/or TZD, with or without other AHMs Note: The body weight effect of SGLT-2 inhibitor may be offset by SU/TZD, and vice-versa, when used in combination. OR • All other monotherapy, such as glinides and AGIs or combination therapy, that does not fall under stratum 1 or 2

Abbreviations: AGI = alpha-glucosidase inhibitor; AHM = antihyperglycemic medication; SGLT = sodium-glucose cotransporter; SU = sulfonylurea; TZD = thiazolidinedione.



Abbreviations: QD = once daily; T = telehealth visit.

Figure GZGQ.1.1. Illustration of study design for clinical protocol J2A-MC-GZGQ.

2. Statistical Hypotheses

The null hypotheses corresponding to the primary objective are as follows:

- $H_{1,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the mean percent change from baseline in body weight at Week 72.
- $H_{2,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the mean percent change from baseline in body weight at Week 72.
- $H_{3,0}$: No difference in 6 mg orforglipron compared to placebo with respect to the mean percent change from baseline in body weight at Week 72.

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

- $H_{4,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 5% or more body weight reduction at Week 72.
- $H_{5,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 5% or more body weight reduction at Week 72.
- $H_{6,0}$: No difference in 6 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 5% or more body weight reduction at Week 72.
- $H_{7,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 10% or more body weight reduction at Week 72.
- $H_{8,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 10% or more body weight reduction at Week 72.
- $H_{9,0}$: No difference in 6 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 10% or more body weight reduction at Week 72.
- $H_{10,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 15% or more body weight reduction at Week 72.
- $H_{11,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 15% or more body weight reduction at Week 72.
- $H_{12,0}$: No difference in 36 mg orforglipron compared to placebo with respect to mean change from baseline in waist circumference at Week 72.
- $H_{13,0}$: No difference in 36 mg orforglipron compared to placebo with respect to mean change from baseline in HbA1c at Week 72.
- $H_{14,0}$: No difference in 12 mg orforglipron compared to placebo with respect to mean change from baseline in HbA1c at Week 72.
- $H_{15,0}$: No difference in 6 mg orforglipron compared to placebo with respect to mean change from baseline in HbA1c at Week 72.
- $H_{16,0}$: No difference in 36 mg orforglipron compared to placebo with respect to mean change from baseline in fasting glucose at Week 72.
- $H_{17,0}$: No difference in 12 mg orforglipron compared to placebo with respect to mean change from baseline in fasting glucose at Week 72.
- $H_{18,0}$: No difference in 6 mg orforglipron compared to placebo with respect to mean change from baseline in fasting glucose at Week 72.

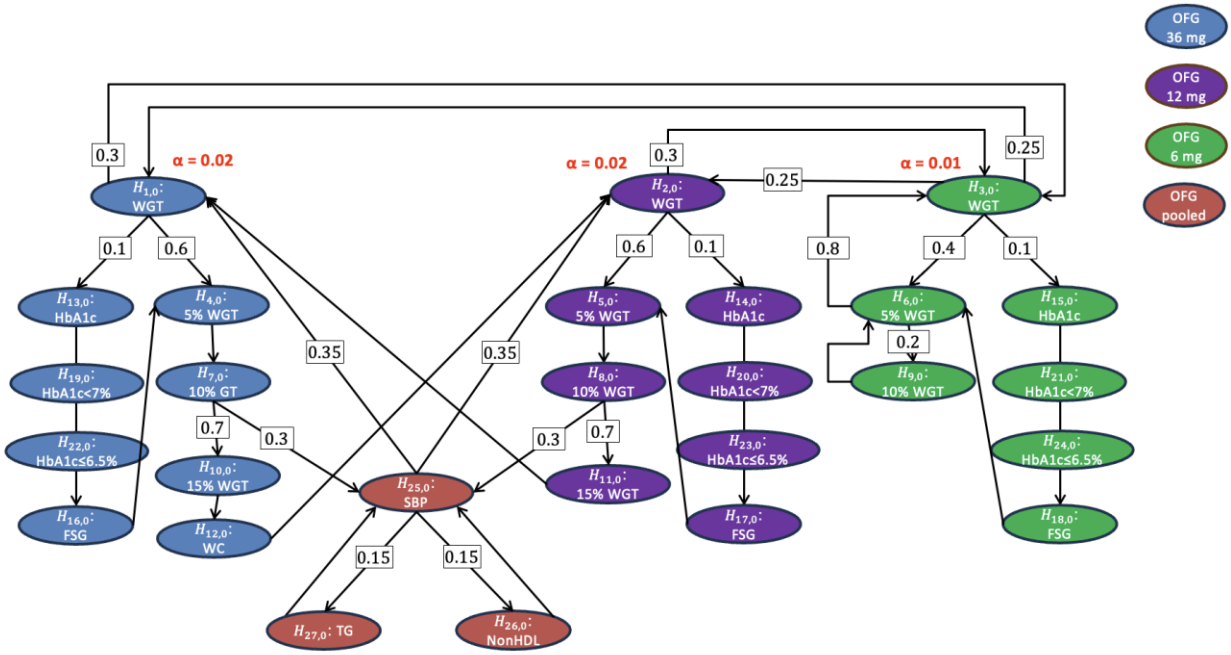
Approved on 26 Jun 2025 GMT

- $H_{19,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants achieving HbA1c target of <7.0% at Week 72.
- $H_{20,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants achieving HbA1c target of <7.0% at Week 72.
- $H_{21,0}$: No difference in 6 mg orforglipron compared to placebo with respect to the percentage of participants achieving HbA1c target of <7.0% at Week 72.
- $H_{22,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants achieving HbA1c target of ≤6.5% at Week 72.
- $H_{23,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants achieving HbA1c target of ≤6.5% at Week 72.
- $H_{24,0}$: No difference in 6 mg orforglipron compared to placebo with respect to the percentage of participants achieving HbA1c target of ≤6.5% at Week 72.
- $H_{25,0}$: No difference in 6 mg, 12 mg, and 36 mg orforglipron pooled compared to placebo with respect to mean change from baseline in systolic BP at Week 72.
- $H_{26,0}$: No difference in 6 mg, 12 mg, and 36 mg orforglipron pooled compared to placebo with respect to mean percent change from baseline in non–high-density lipoprotein cholesterol at Week 72.
- $H_{27,0}$: No difference in 6 mg, 12 mg, and 36 mg orforglipron pooled compared to placebo with respect to mean percent change from baseline in triglycerides at Week 72.

2.1. Multiplicity Adjustment

Multiplicity adjusted analyses will be performed on the primary and key secondary objectives to control the overall family-wise Type 1 error rate at a 2-sided alpha level of 0.05. The graphical multiple testing procedure described in Bretz and colleagues (2009, 2011) will be used. This approach is a closed testing procedure; hence, it strongly controls the family-wise Type 1 error rate across all hypotheses (Alosh et al. 2014).

Figure GZGQ.2.1 illustrates the final graphical testing scheme, including testing order, interrelationships, Type 1 error allocation for the primary and key secondary objectives, and the associated propagation. The treatment regimen estimand will be the primary estimand, with the efficacy estimand considered supportive. Since these 2 types of estimands are intended for distinct purposes, no multiplicity adjustment will be made for conducting separate analyses on the same objectives. Unless otherwise specified, there will be no adjustment for multiple comparisons for any other analyses outside the primary and key secondary objectives.



Abbreviation: OFG = orforglipron.

Figure GZGQ.2.1. Graphical testing procedure.

3. Analysis Sets

For the purpose of analysis, the analysis populations are defined in this table along with their attributes.

Table GZGQ.3.1. Participant Analysis Sets

Participant Analysis Set	Description
Entered participants	**Definition:** All participants who sign informed consent[a]. **Purpose:** Used for providing summaries for screen failures and the associated reasons. **Treatment Groups:** None **Inferential Comparisons:** None
Randomized participants	**Definition:** All participants who are randomly assigned a study intervention. **Purpose:** Used for listings of disposition and treatment assignment summaries of disposition, demographics, historical illness, preexisting conditions, and analyses of efficacy and health outcomes. **Treatment Groups:** OFG 6 mg, OFG 12 mg, OFG 36 mg, PBO **Inferential Comparisons:** OFG 6 mg vs. PBO; OFG 12 mg vs. PBO; OFG 36 mg vs. PBO; OFG pooled vs. PBO
Safety participants	**Definition:** All participants who are randomly assigned a study intervention and who take at least 1 dose of study intervention. **Purpose:** Used for all safety analyses. **Treatment Groups:** OFG 6 mg, OFG 12 mg, OFG 36 mg, PBO **Inferential Comparisons:** OFG 6 mg vs. PBO; OFG 12 mg vs. PBO; OFG 36 mg vs. PBO

Abbreviations: OFG = orforglipron; PBO = placebo.

[a] Refers to the informed consent for the study.

The following data points sets are defined for all parameters except for the glycemic-related parameters.

Table GZGQ.3.2. Data Points Sets for Non–Glycemic-Related Parameters

Data Points Sets	Description
Treatment regimen estimand data points set	All data points obtained during the treatment period defined as at or after baseline and up to the last visit within the treatment period, regardless of study intervention discontinuation or initiation of prohibited weight management treatments. Baseline is defined in Section 4.1.1.
Efficacy estimand data points set	All data points obtained during the treatment period defined as at or after baseline and up to the earliest date of study intervention discontinuation or initiation of prohibited weight management treatments. Baseline is defined in Section 4.1.1 and the prohibited weight management treatments are specified in Section 6.3 and Table GZGQ.6.3.

The following data points sets are defined for all glycemic-related parameters. The parameters in scope include fasting glucose and HbA1c.

Table GZGQ.3.3. Data Points Sets for Glycemic-Related Parameters

Data Points Sets	Description
Treatment regimen estimand data points set	All data points obtained during the treatment period defined as at or after baseline and up to the last visit within the treatment period, regardless of study intervention discontinuation or initiation of prohibited weight management treatments or glycemic rescue therapy. Baseline is defined in Section 4.1.1.
Efficacy estimand data points set	All data points obtained during the treatment period defined as at or after baseline and up to the earliest date of discontinuation of study intervention or initiation of prohibited weight management treatments or glycemic rescue therapy. Baseline is defined in Section 4.1.1 and the prohibited weight management treatments, glycemic rescue therapies and prohibited glycemic therapies are specified in Section 6.3 and Table GZGQ.6.3.

The following data points set is defined for the safety analyses.

Table GZGQ.3.4. Data Points Sets for Safety Analyses

Data Points Sets	Description
Safety data points set	All data points obtained during the treatment period and the follow-up period defined as at or after baseline and up to the date of study withdrawal or study completion, including the follow-up period, regardless of intercurrent events. Baseline is defined in Section 4.1.1.

4. Statistical Analyses

4.1. General Considerations

Statistical analysis of this study will be the responsibility of 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 this SAP or the clinical study report (CSR). Additional exploratory data analyses may be conducted as deemed appropriate.

Details about the analyses regarding demographic and baseline characteristics, historical illnesses and preexisting conditions, treatment compliance, concomitant medications, and important protocol deviations can be found in Appendices 1 through 5 (Section 6.1 through Section 6.5), respectively.

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

Unless otherwise noted, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and the confidence interval (CI) will be calculated at 95%, 2-sided.

Change from baseline will be calculated as the value of interest at the visit minus the baseline value. In general, percent change from baseline will be calculated as the value of change from baseline divided by the baseline value in 100% scale. For some specific predefined variables, percent change from baseline might be calculated using the log transformation. If baseline value is missing for a particular variable, then the change from baseline and percent change from baseline will not be calculated when deriving the summary statistics.

For stratification factors at baseline, a strata variable is defined for statistical modeling to consist of 6 joint levels. The strata variable includes sex (female, male) and background oral AHMs (classified according to potential effect on body weight: weight loss, weight gain, weight neutral and others). Although, country is also a stratification factor, geographic region (North America, Europe, Asia, Australia or Central/South America) will be included in statistical modeling as a separate factor to improve model convergence.

Data may exist at postbaseline visits where the variable was not scheduled to be collected, for example, early discontinuation visits. In these situations, data from unscheduled visits and from the early discontinuation visit that do not correspond to the planned collection schedule will be excluded from the mixed model for repeated measures (MMRM), analysis of covariance (ANCOVA), or logistic regression analysis, unless otherwise specified (Andersen and Millen 2013).

Handling of missing data is addressed in Section 4.1.2 and Table GZGQ.4.6. Section 3 provides definitions for the participant analysis set and data points set which includes definitions for censored data where appropriate. Handling spurious data is addressed in Section 6.5 for important protocol deviations and Protocol Section 10.1.7 for data quality assurance.

4.1.1. Definition of Baseline

Unless otherwise specified, baseline for efficacy assessments is defined as the last available non-missing measurement prior to the first dose of study intervention, which in most cases will be the measurement recorded at Week 0 (Visit 3). If there are no doses of study intervention, baseline will be defined as the last available non-missing measure on or prior to randomization. In cases where the measurement is taken on the same day (where time is not collected or not reliable) as the first dose, this measurement will be used as the baseline values for data analysis. For patient-reported outcome measures, data obtained at Visit 3, regardless of the timing relative to first dose, will serve as baseline.

For the safety assessments, the definition of baseline and postbaseline are specified in Table GZGQ.4.1.

Table GZGQ.4.1. Baseline and Postbaseline Definitions for Safety Analyses

Analysis Type	Baseline Period	Postbaseline Period
TEAEs	Starts from informed consent date and ends prior to the first dose (typically at Week 0).	Starts after the first dose of study intervention and ends at the end of the follow-up period or the date of study withdrawal[a].
TE abnormal laboratory values, vital signs, and ECGs	Starts from informed consent date and ends prior to the first dose (typically at Week 0). All scheduled and unscheduled measurements will be included.	Starts after the first dose and ends at the end of the follow-up period or the date of study withdrawal. All scheduled and unscheduled measurements will be included.
Change from last baseline to each postbaseline week and to last postbaseline for laboratory values, vital signs, and ECGs	When "last baseline" is used, starts from informed consent date and ends prior to the first dose (typically at Week 0).	Starts after the first dose and ends at the end of the safety follow-up period or the date of study withdrawal. Only scheduled visits and early termination visits that fall on the scheduled visits will be included.

Abbreviations: ECG = electrocardiogram; TE = treatment-emergent; TEAE = treatment-emergent adverse event.

[a] For events occurring on the day of first dose, the case report form collected information will be used to determine whether the event was pre- versus post-treatment if available. If the relevant information is not available, then the events will be counted as postbaseline.

4.1.2. Analysis Methods

The analysis methods are consistent with the desired estimands and the FDA guidance on "Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products Guidance Document" (FDA 2023).

4.1.2.1. Analysis Methods for Treatment Regimen Estimand

4.1.2.1.1. Method for Continuous Variables

The ANCOVA model will be used to analyze continuous measurements at Week 72 and will be guided by treatment regimen estimand. The model will include

- Treatment group as a factor variable
- Region as a factor variable
- Strata as a factor variable
- Baseline value (of the dependent variable)
- Interaction between strata variable and treatment group
- Interaction between baseline value and treatment group

The estimated treatment group effect and comparison between 6 mg orforglipron, 12 mg orforglipron, or 36 mg orforglipron and placebo will be reported together with the variability estimated using the robust inference (Ye et al. 2022; FDA 2023). The associated 2-sided 95% CI and corresponding p-values will also be reported. If the model fails to converge, all the interaction terms will be removed before the model fitting. The addition of interaction terms is not intended to estimate the heterogeneity effect but to provide robustness and efficiency for the estimate of treatment comparisons on the unconditional effect. 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 GZGQ.4.2.

For some variables, both the baseline and the postbaseline values will be log-transformed before fitting the ANCOVA. The treatment group estimates will be the percent change from baseline while the treatment contrasts will be the relative change in orforglipron 6 mg, orforglipron 12 mg, or orforglipron 36 mg compared to the placebo (%). In these cases, least squares (LS) means 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.

A linear contrast, averaging estimates from the individual doses, will be used to estimate the treatment effect of the pooled doses compared with placebo.

4.1.2.1.2. Primary Multiple Imputation Strategy

Discontinuation is expected to be uncommon. Participants who discontinue the study intervention (that is, discontinue study treatment) will be encouraged to continue in the study 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 the best precautions, missing data should be imputed in a fashion 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 (Week 72 visit) shown in Figure GZGQ.4.1.

Approved on 26 Jun 2025 GMT

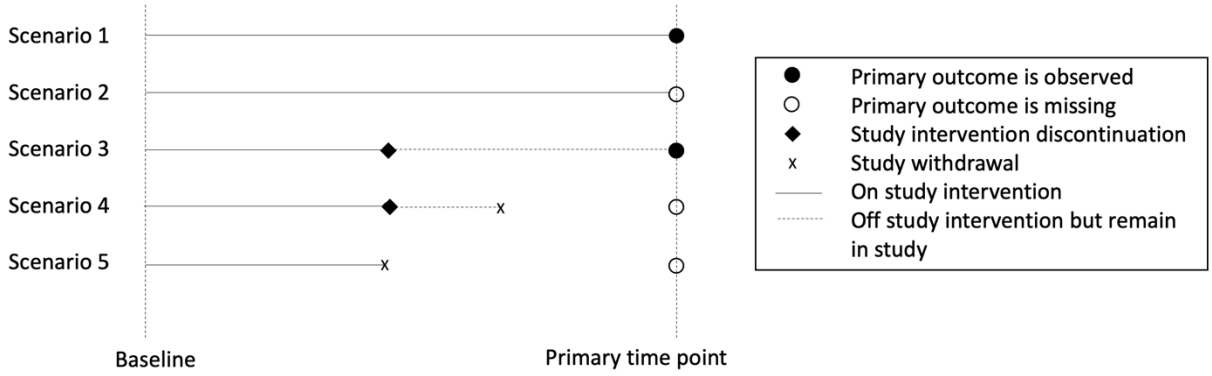


Figure GZGQ.4.1. Treatment journey of a trial participant.

Scenario 5 covers the situation 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. Additional details on scenario 5 are provided in Table GZGQ.4.3.

In principle, missing data due to permanent discontinuation of study intervention (Scenario 4 and 5A) will be imputed by treatment group using retrieved dropouts (multiple imputation retrieved dropouts [MI-RD]), namely using multiple imputation based on data retrieved from participants who permanently discontinued study intervention but continued in the study 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, model implemented does not converge), an alternative multiple imputation method (placebo-washout method) (Wang 2023) will be used. Any missing values at baseline will be imputed in the same model when imputing the endpoint.

Table GZGQ.4.2. Imputation Procedure

Scenario (Study Intervention/Treatment Period Discontinuation; Missingness Status 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 postbaseline visits, across all time points from Scenarios 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 treatment period 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 the endpoint visit data from Scenario 3 (MI-RD) 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 treatment period discontinuation reasons classified as Category 5A in Table GZGQ.4.3), 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 treatment period discontinuation reasons classified as Category 5B in Table GZGQ.4.3), 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 retrieved dropouts; N/A = not applicable.

Table GZGQ.4.3. Categorization of Treatment Period Discontinuation Reasons

Disposition Reason	Associated Subcategories	Category
Adverse Event		5A
Death		5A
Protocol Deviation		5A
Pregnancy		5A
Noncompliance With Study Drug		5A
Lack of Efficacy	Investigator	5A
	Study subject	5A
Withdrawal by Subject	Concern about study procedures/perceived risks	5A
	Health insurance changes	5A
	Scheduling conflicts	5A
	Subject is moving or has moved	5A
	Personal issue unrelated to trial	5A
	Due to epidemic/pandemic	5B
	Other	5A
Physician Decision	Due to epidemic/pandemic	5B
	Other	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]).

4.1.2.1.3. Modified Multiple Imputation Strategy

A modified multiple imputation analysis is planned as a sensitivity analysis to assess the robustness of the primary efficacy results. An additional analysis will be conducted and data observed from Scenario 3 will be used to impute Scenario 5B.

4.1.2.1.4. Multiple Imputation–Based Tipping-Point Analysis

A multiple imputation–based tipping-point (MI-TP) analysis is planned as a sensitivity analysis to explore how different patterns of percent body weight change post treatment period discontinuation by different treatment groups could impact the treatment comparisons.

To start with, missing data are imputed according to the primary multiple imputation (PMI), as shown in Section 4.1.2.1.2. A penalty is then added to those imputed values at Week 72 visit. The analysis varies the magnitude of the penalties added for both treatment groups under comparison and evaluates the impact these would have on the study conclusion. A 2-dimentional space of penalties will be assessed for orforglipron 36 mg, orforglipron 12 mg, or orforglipron 6 mg and placebo. MI-TP aims to evaluate the robustness of the superiority claim to the assumptions of using the observed data to impute the missing body weight in all treatment groups.

4.1.2.1.5. Method for Binary Variables

The binary outcomes will be analyzed with following the procedure:

- Using the same imputation strategy – PMI as in Section 4.1.2.1.2 on the underlying continuous endpoint that determines the binary value. For example, the continuous body weight value at Week 72 visit will be imputed first.
- Transform the observed and imputed continuous value to the binary value, for example, convert the continuous body weight value at Week 72 visit to whether the corresponding percent change from baseline meets a certain threshold (Ma et al. 2022).
- Fit a logistic regression model to the data with the following terms:
    - Treatment group as a factor variable
    - Region as a factor variable
    - Strata as a factor variable
    - Baseline value (of the dependent variable)
    - Interaction between treatment group and strata, and
    - Interaction between treatment group and baseline value.
- Provide estimates and inferences of unconditional treatment effects defined by the risk difference and/or relative risk based on the delta-method using the formula provided (Ye et al. 2023).
- Derive the final inference using Rubin's Rule, which combines estimates from multiple imputed datasets, with the imputation procedure for creating a single imputed dataset detailed in Table GZGQ.4.2.

The statistical testing for risk difference will be used as the primary approach for treatment comparisons.

4.1.2.2. Analysis Methods for Efficacy Estimand

4.1.2.2.1. Method for Continuous Variables

A maximum-likelihood based MMRM will be used to analyze continuous measurements at each postbaseline visit guided by efficacy estimand. Missing data should be minimized at the best precaution. The hypothetical strategy will be used to handle the ICEs. For MMRM analysis, only data collected before the occurrence of any ICEs will be used in the MMRM analysis. Through the MMRM, the potential efficacy measures (after the ICEs) will be implicitly imputed as if participants did not have ICEs. Participant dose modification and interruption will not disqualify their measurements from being included into the model.

The MMRM model will include the following terms:

- Treatment group as a factor variable
- Visit as a factor variable
- Region as a factor variable
- Strata as a factor variable
- Baseline value (of the dependent variable)
- Interaction between treatment group, strata variable and visit.

Approved on 26 Jun 2025 GMT

- Interaction between treatment group, baseline value and visit.

The estimated treatment group effect and comparison between 6 mg orforglipron, 12 mg orforglipron, or 36 mg orforglipron and placebo at the scheduled visits will be reported together with the variability estimated using the robust inference (Wang and Du 2021). The sandwich estimator (Diggle et al. 1994) for the variance-covariance matrix will be used. The addition of 3-way interaction terms is not intended to estimate the heterogeneity effect but to provide robustness and efficiency for the estimate of treatment comparisons on the unconditional effect. The associated 2-sided 95% CI and corresponding p-values will also be reported. An unstructured covariance matrix by each treatment group will be used to model the within-participant errors, assuming heteroscedasticity and the measurements for different participants are independent. If the model fails to converge, the model is simplified to include the following terms removing 3-way interactions and the heteroscedasticity assumption:

- Treatment group as a factor variable
- Visit as a factor variable
- Region as a factor variable
- Strata as a factor variable
- Baseline value (of the dependent variable)
- Interaction between treatment group and visit
- Interaction between strata variable and visit
- Interaction between baseline value and visit

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

- Heterogeneous Toeplitz
- Heterogeneous autoregressive
- Heterogeneous compound symmetry
- Homogeneous Toeplitz
- Homogeneous autoregressive
- Homogeneous compound symmetry

The first covariance structure that converges will be used.

For some variables, both the postbaseline response variables and baseline variable will be log-transformed before fitting the MMRM. The treatment group estimates will be the percent change from baseline while the treatment contrasts will be the relative change in orforglipron 6 mg, orforglipron 12 mg, or orforglipron 36 mg compared to the placebo (%) over the scheduled visits.

A linear contrast, averaging estimates from the individual doses, will be used to estimate the treatment effect of the pooled doses compared with placebo.

4.1.2.2.2. Method for Binary Variables

The binary outcomes will be analyzed with the following procedure:

- Impute the missing continuous-valued measurements at the scheduled visits using the randomized participants with efficacy estimand data points set assuming MAR with multiple imputation.
- At the visit of interest, transform the observed and imputed continuous value to the binary value, for example, convert the continuous body weight value at a visit to whether the corresponding percent change from baseline meets a certain threshold.
- Fit a logistic regression model to the transformed binary data with the following terms:
    - Treatment group as a factor variable
    - Region as a factor variable
    - Strata as a factor variable
    - Continuous baseline value (of the dependent variable)
    - Interaction between treatment group and strata, and
    - Interaction between treatment group and baseline value
- Provide for the visit of interest the estimates and inferences of unconditional treatment effects defined by the risk difference and/or risk ratio based on the delta-method using the formula provided (Ye et al. 2023).
- Derive the final inference using Rubin's Rule by combining estimates from multiple imputed datasets.

The statistical testing for risk difference will be considered the primary method for treatment comparisons.

4.1.2.3. Analysis Methods for Safety

In general, safety assessments will be guided by an estimand comparing safety of orforglipron doses with placebo regardless of intercurrent events. Thus, safety analyses will be conducted using the safety participants based on data observed during treatment period plus follow-up period, that is, the period from first dose of treatment to end of follow-up visit or the date of study withdrawal. It is noted that for some safety endpoints, such as, study intervention discontinuation due to adverse event (AE), and so on, the analyses will be based on data observed during treatment period only.

Fisher's exact test will be used for treatment comparisons of percentages. The risk differences and 95% CIs will be provided.

For selected continuous safety measures, unless otherwise specified, treatment group differences of mean change or percent mean change from baseline at all scheduled visits will be assessed via an MMRM using maximum likelihood, which will include the following terms:

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

If the data does not warrant the MMRM model, then an ANCOVA model with treatment group as a fixed effect and the continuous baseline value as a covariate will be used.

No explicit imputation will be conducted for safety measures. Some parameters (such as urinary albumin to creatinine ratio [UACR], p-amylase, and lipase) may be log-transformed before fitting the MMRM as specified in Section 4.6. In these cases, LS means 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 of orforglipron treatment groups compared to placebo in percent change.

The Kaplan-Meier product limit method will be used to estimate the cumulative event-free survival rates over time for the time to event analyses. An unstratified Cox proportional hazards regression analysis will be used to compare hazard rates among treatments. A unstratified log-rank test will be used to calculate p-values. Time-to-event for the specific safety event of interest will be calculated from the date of first dose.

The logistic regression model will be used for hypoglycemia incidence for treatment comparisons for the overall postbaseline period. The model will include the fixed effects of treatment. Where necessary, the rate of events will be analyzed using a generalized linear mixed-effects model assuming the number of events follow a negative binomial distribution with the mean modeled using treatment as a fixed effect for the overall postbaseline period. The logarithm of days during analysis interval will be adjusted as an offset to account for possible unequal duration of follow-up between participants. When the negative binomial model fails to converge, and/or the number of events from at least one treatment group is less than or equal to 10, an empirical method for statistical comparison will be applied.

Some safety analyses may be conducted after excluding data after the initiation of new AHMs.

4.2. Participant Dispositions

The participant dispositions for the screening period, the study intervention, the treatment period and/or the follow-up period will be collected in CRFs with corresponding primary reason. The study completion for a participant is defined as the participant completing both the treatment period and the follow-up period, regardless of completion of study intervention.

The planned listings and summary tables for dispositions are provided in Table GZGQ.4.4. No inferential analysis will be performed.

Table GZGQ.4.4. Listings and Summary Tables Related to Dispositions

Analysis	Population/Period
Summary of disposition (prior to randomization)	Entered participants
Patient allocation by region, country, and center/site	Entered participants
Summary of study and study intervention disposition	Randomized participants/TP + FP (TP for study intervention disposition)
Kaplan-Meier plot of time to study discontinuation	Randomized participants/TP + FP

Kaplan-Meier plot of time to study intervention discontinuation	Safety participants/TP
Kaplan-Meier plot of time to study intervention discontinuation due to AEs	Safety participants/TP
Listing of randomization	Randomized participants
Listing of randomized participants who were discontinued from the study intervention due to inadvertent enrollment	Randomized participants
Listing of study and study intervention disposition	Randomized participants

Abbreviations: AE = adverse event; FP = follow-up period; TP = treatment period.

Time-to-study intervention discontinuation will be calculated from the date of first dose. Time-to-study discontinuation will be calculated from the date of randomization.

4.3. Primary Endpoint/Estimand Analysis

4.3.1. Definition of Endpoint(s)

The primary efficacy endpoint is percent change in body weight at 72 weeks.

As specified in Table GZGQ.4.5, the percent change in body weight is defined as

*(postbaseline body weight [kg] – baseline body weight [kg]) / baseline body weight [kg] * 100%.*

4.3.2. Sensitivity Analyses

The sensitivity analyses are specified in Section 4.1.2.1.3 and Section 4.1.2.1.4 (MI-TP).

4.3.3. Main Analytical Approach

The analytical approaches are specified in Section 4.1.2.1.1 and Section 4.1.2.1.2 (ANCOVA with PMI) and Section 4.1.2.2.1 (MMRM).

4.3.4. Supplementary Analyses

No planned supplementary analysis for primary endpoint.

4.4. Secondary Endpoints/Estimands Analysis

Table GZGQ.4.5 includes the description and derivation of the primary and secondary endpoints for the efficacy measures and patient-reported outcomes.

Table GZGQ.4.6 provides the detailed analyses including endpoint, estimand, data points set, method and imputation, population, and time point for the primary and secondary endpoints.

Table GZGQ.4.5. Description and Derivation of Efficacy Measures and Patient Reported Outcomes

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
Body weight	Per Protocol Section 10.9: • Body weight measurements should be done in a consistent manner using a calibrated electronic scale capable of measuring weight in kilograms to 1 decimal place. • All weights for a given participant should be measured using the same scale, whenever possible, at approximately the same time in the morning after evacuation of bladder contents. • Body weight will be measured in fasting state at all visits. If the participant is not fasting, the participant should be called in for a new visit within the visit window to have the fasting body weight measured.	Body weight (kg)	As measured	Single item, missing if missing
		Change from baseline in body weight (kg)	Calculated as postbaseline body weight (kg) − baseline body weight (kg)	Missing if baseline or postbaseline value is missing
		Percent change from baseline in body weight (%)	Calculated as (postbaseline body weight [kg] − baseline body weight [kg]) / baseline body weight [kg] * 100 (%)	Missing if baseline or postbaseline value is missing
		≥x% body weight reduction from baseline where, x = 5, 10, 15, 20	Response = yes if at least a x% reduction in body weight from baseline, that is, percent change from baseline in body weight ≤ −x%	Missing if baseline or postbaseline value is missing
BMI	BMI: Round to 1 decimal point. For example, a BMI of 26.6 kg/m^2 should **not** be rounded to 27.0 kg/m^2.	BMI (kg/m^2)	BMI will be calculated as: Weight (kg) / (height [m])^2	Missing if weight or height is missing
		Change from baseline in BMI (kg/m^2)	Calculated as postbaseline BMI (kg/m^2) − baseline BMI (kg/m^2)	Missing if baseline or postbaseline value is missing
		BMI threshold value of <x kg/m^2, where x = 25, 27, 30, 35	Response = yes if the postbaseline BMI value <x kg/m^2	Missing if postbaseline value is missing
Waist circumference	Per Protocol Section 10.9: • Waist circumference should be measured in the horizontal plane and at the midpoint between the lower margin of the last palpable rib and	Waist circumference (cm)	The average of 2 measures	If there is at least one measurement, take the average of all nonmissing measurements; otherwise, set it to

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
	the top of the iliac crest. • Measurements should be taken at the end of a normal expiration using a nonstretchable measuring tape. The tape should lie flat against the skin without compressing the soft tissue. • The waist circumference should be measured twice, rounded to the nearest 0.5 cm. The measuring tape should be removed between the 2 measurements. Both measurements will be recorded in the CRF. If the difference between the 2 measurements exceeds 1 cm, this set of measurements should be discarded and the 2 measurements repeated.			be missing
		Change from baseline in waist circumference (cm)	Calculated as postbaseline Waist circumference (cm) − baseline Waist circumference (cm)	Missing if baseline or postbaseline value is missing
Blood pressure	Per Protocol Section 10.9: • Have the participant sit quietly for about 5 minutes before vital signs measurements are taken. • Take 3 measurements from the same arm, preferably the nondominant arm. • Measure the recordings at least 1 minute apart. • Blood pressure must be taken with an automated blood pressure instrument	SBP (mmHg)	The average of 3 measures	If there is at least 1 measurement, take the average of all nonmissing measurements; otherwise, set it to be missing
		Change from baseline in SBP (mmHg)	Calculated as postbaseline SBP (mmHg) − baseline SBP (mmHg)	Missing if baseline or postbaseline value is missing
		DBP (mmHg)	The average of 3 measures	If there is at least 1 measurement, take the average of all nonmissing measurements; otherwise, set it to be missing
		Change from baseline in DBP (mmHg)	Calculated as postbaseline DBP (mmHg) − baseline DBP (mmHg)	Missing if baseline or postbaseline value is missing

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
Lipid parameter	HDL-cholesterol, triglycerides, and total cholesterol will be assayed by Lilly designated laboratory. LDL-cholesterol, non–HDL-cholesterol, and VLDL-cholesterol will be calculated by Lilly designated laboratory.	Triglycerides	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in triglycerides (%)	Calculated as log(postbaseline triglycerides) − log (baseline triglycerides) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		Total cholesterol	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in total cholesterol (%)	Calculated as log(postbaseline total cholesterol) − log(baseline total cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		Non–HDL-cholesterol	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in non–HDL-cholesterol (%)	Calculated as log(postbaseline non–HDL-cholesterol) − log(baseline non–HDL-cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		LDL-cholesterol	As provided If triglycerides are >400 mg/dL, the direct LDL will be directly measured Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in LDL-cholesterol (%)	Calculated as log(postbaseline LDL-cholesterol) − log(baseline LDL-cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		HDL-cholesterol	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in HDL-cholesterol (%)	Calculated as log(postbaseline HDL-cholesterol) − log(baseline HDL-cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		VLDL-cholesterol	As provided Log transformation before the analysis	Single item, missing if missing

Approved on 26 Jun 2025 GMT

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
		Percent change from baseline in VLDL-cholesterol (%)	Calculated as log(postbaseline VLDL-cholesterol − log(baseline VLDL-cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
Glycemic control	Fasting glucose will be assayed by Lilly-designated laboratory.	Fasting glucose (mg/dL and mmol/L)	As provided	Single item, missing if missing
		Change from baseline fasting glucose (mg/dL and mmol/L)	Calculated as Postbaseline fasting glucose − Baseline fasting glucose	Missing if baseline or postbaseline value is missing
	HbA1c will be assayed by Lilly-designated laboratory.	HbA1c (% and mmol/mol)	As provided	Single item, missing if missing
		Change from baseline HbA1c (% and mmol/mol)	Calculated as Postbaseline HbA1c − Baseline HbA1c	Missing if baseline or postbaseline value is missing
		HbA1c target value of ≤6.5%	Response = yes if the postbaseline HbA1c value ≤6.5%	Missing if postbaseline value is missing
		HbA1c target value of <7%	Response = yes if the postbaseline HbA1c value <7%	Missing if postbaseline value is missing
	Participants must collect two 7-point SMBGs on 2 nonconsecutive days before assigned visits. A 7-point SMBG consists of measurements before and 2 hours after each of 3 main meals within the same day and at bedtime, that is • Point 1: before morning meal fasting • Point 2: 2 hours after morning meal • Point 3: before midday meal • Point 4: 2 hours after midday meal	Point x SMBG (mg/dL) x = 1, 2, …,7	The average of 2 measures in the 2 nonconsecutive days at each time point	If there is at least 1 measurement, take the average of all nonmissing measurements; otherwise, set it to be missing
		Change from baseline in point x SMBG (mg/dL)	Calculated as postbaseline point x SMBG (mg/dL) − baseline point x SMBG (mg/dL)	Missing if baseline or postbaseline value is missing
		Mean of all 7-point SMBG (mg/dL)	Calculated as (average of point 1 + average of point 2 +… + average of point 7)/7	If there is at least 4 measurements, take the average of

Approved on 26 Jun 2025 GMT

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
	• Point 5: before evening meal • Point 6: 2 hours after evening meal • Point 7: bedtime The participant needs to collect these measurements within 14 days before the assigned visits. If more than two 7-point SMBG profiles are available, the 2 most recent nonconsecutive profiles should be used. For SMBG, valid SMBG profiles will be used for analysis, defined as having non-missing values at ≥4 time points (including at least one post-meal point) among the 7 pre-specified time points and being collected within 2 weeks prior to the given visit.			all nonmissing measurements; otherwise, set it to be missing
		Change from baseline in mean of all 7-point SMBG (mg/dL)	Calculated as Postbaseline mean of all 7-point SMBG (mg/dL) − baseline mean of all 7-point SMBG (mg/dL)	Missing if baseline or postbaseline value is missing
		Mean of premeal SMBG (mg/dL)	Calculated as (average of point 1 + average of point 3 + average of point 5)/3	If there is at least 1 measurement, take the average of all nonmissing measurements; otherwise, set it to be missing
		Change from baseline in mean of premeal SMBG (mg/dL)	Calculated as postbaseline mean of premeal SMBG (mg/dL) − baseline mean of premeal SMBG (mg/dL)	Missing if baseline or postbaseline value is missing
		Mean of 2-hour postmeal SMBG (mg/dL)	Calculated as (average of point 2 + average of point 4 + average of point 6)/3	If there is at least 1 measurement, take the average of all nonmissing measurements; otherwise, set it to be missing
		Change from baseline in postmeal SMBG (mg/dL)	Calculated as Postbaseline mean of postmeal SMBG (mg/dL) − baseline mean of postmeal SMBG (mg/dL)	Missing if baseline or postbaseline value is missing
Fasting Insulin	Per Protocol Section 10.2: Fasting insulin will be assayed by Lilly-designated laboratory.	Fasting insulin	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in fasting insulin (%)	Calculated as log(postbaseline fasting insulin) − log(baseline fasting insulin) then will transform back to percent change	Missing if baseline or postbaseline value is missing
Renal function	Per Protocol Section 10.2:	UACR	As provided	Single item,

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
	UACR and eGFR will be assayed by Lilly-designated laboratory.		Log transformation before the analysis	missing if missing
		Percent change from baseline in UACR (%)	Calculated as log(postbaseline UACR) − log(baseline UACR) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		eGFR CKD-EPI Cystatin C (ml/min/1.73m^2)	As provided as: eGFR (CKD-EPI Cystatin C 2012)	Single item, missing if missing
		Change from baseline in eGFR CKD-EPI Cystatin C (ml/min/1.73m^2)	Calculated as postbaseline eGFR CKD-EPI Cystatin C (ml/min/1.73m^2) − Baseline eGFR CKD-EPI Cystatin C (ml/min/1.73m^2)	Missing if baseline or postbaseline value is missing
		eGFR CKD-EPI Creatinine (ml/min/1.73m^2)	As provided as: eGFR (CKD-EPI Creatinine 2021)	Single item, missing if missing
		Change from baseline in eGFR CKD-EPI Creatinine (ml/min/1.73m^2)	Calculated as: postbaseline eGFR CKD-EPI Creatinine (ml/min/1.73m^2) − Baseline eGFR CKD-EPI Creatinine (ml/min/1.73m^2)	Missing if baseline or postbaseline value is missing
Inflammatory biomarker	Per Protocol Section 10.2: The hsCRP will be assayed by Lilly-designated laboratory.	hsCRP (mg/L)	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in hsCRP (%)	Calculated as log(postbaseline hsCRP (mg/L)) − log(baseline hsCRP (mg/L)) then will transform back to percent change	Missing if baseline or postbaseline value is missing
FFAs	Per Protocol Section 10.2: FFAs will be performed by the central laboratory and will be assayed by Lilly-designated laboratory.	FFAs (mmol/L)	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in FFAs (%)	Calculated as log(postbaseline FFAs (mmol/L)) − log(baseline FFAs (mmol/L)) then will transform back to percent change	Missing if baseline or postbaseline value is missing
HOMA	The HOMA2 B estimates steady state beta cell function.	HOMA2 B (insulin)	Calculated from the link[a] using fasting glucose and insulin laboratory values	Missing if fasting glucose or insulin is missing

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
		Percent change from baseline in HOMA2 B (insulin)(%)	Calculated as log[postbaseline HOMA2 B (insulin)] – log[baseline HOMA2 B (insulin)]	Missing if baseline or postbaseline value is missing
		HOMA2 B (C-peptide)	Calculated from the link[a] using fasting glucose and C-peptide laboratory values	Missing if fasting glucose or c-peptide is missing
		Percent change from baseline in HOMA2 B (C-peptide) (%)	Calculated as log[postbaseline HOMA2 B (C-peptide)] – log[baseline HOMA2 B (C-peptide)]	Missing if baseline or postbaseline value is missing
	The HOMA2 IR estimates insulin resistance	HOMA2 IR (insulin)	Calculated from the link[a] using fasting glucose and insulin laboratory values	Missing if fasting glucose or insulin is missing
		Percent change from baseline in HOMA2 IR (insulin) (%)	Calculated as log[postbaseline HOMA2 IR (insulin)] – log[baseline HOMA2 IR (insulin)]	Missing if baseline or postbaseline value is missing
		HOMA2 IR (C-peptide)	Calculated from the link[a] using fasting glucose and C-peptide laboratory values	Missing if fasting glucose or c-peptide is missing
		Percent change from baseline in HOMA2 IR (C-peptide) (%)	Calculated as log[postbaseline HOMA2 IR (C-peptide)] – log[baseline HOMA2 IR (C-peptide)]	Missing if baseline or postbaseline value is missing
SF-36	The Short Form 36 Version 2 Health Survey (SF-36v2) will assess health-related quality of life. The SF-36v2 acute form, 1-week recall version is a 36-item generic, participant-completed measure designed to assess the following 8 domains: • Physical functioning • Role-physical • Bodily pain • General health • Vitality • Social functioning	SF-36 domain scores and SF-36 component scores	Per copyright owner, the Quality Metric Health Outcomes™ Scoring Software will be used to derive SF-36 domain and component scores. After data quality-controls, the SF-36 software will recalibrate the item-level responses for calculation of the domain and component scores. These raw scores will be transformed into the domain scores (T-scores) using the 1-week recall period. This entails exporting the patient data in a CSV or tab-delimited file for import, generation of the SF-36 scores and reports, and export of the calculated scores in a	Missing data handling offered by SF-36 software will be used. "Maximum Data Recovery" will be selected for Missing Score Estimator in the software.

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
	• Role-emotional, and • Mental health. The Physical Functioning domain assesses limitations due to health now. The Physical functioning domain does not have a recall period although the remaining domains assess functioning "in the past week." Each domain is scored individually and information from these 8 domains is further aggregated into 2 health component summary scores, a Physical Component Summary, and a Mental Component Summary. Items are answered on Likert scales of varying lengths (3-point, 5-point, or 6-point scales). Scoring of each domain and both summary scores are norm-based and presented in the form of T-scores, with a mean of 50 and SD of 10. Higher scores indicate better levels of function and/or better health (Maruish 2011).		CSV or tab-delimited file.	
		SF-36 change from baseline for domains and component scores (PCS and MCS)	Calculated as postbaseline SF-36 score − baseline SF-36 score	Missing if baseline or postbaseline value is missing
PGI-S	The PGI-S Physical Function Weight scale is designed to assess the participants' overall perception of their condition. This is a single global item that asks participants to rate how their weight limited their ability to perform physical activities in the past 7 days.	PGI-S rating	As assessed The responses are on a 5-point scale ranging from "not at all limited" to "extremely limited."	Single item, missing if missing
PGI-C	The PGI-C Physical Function Weight scale is designed to assess the participants' overall perception of the efficacy of treatment. This is a single global item that asks	PGI-C rating	As assessed The responses are based on a 5-point scale ranging from "much better" to "much worse."	Single item, missing if missing

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
	participants to rate the overall change in their ability to perform physical activities due to their weight since starting the study intervention.			
IWQOL-Lite-CT	The IWQOL-Lite-CT Version (Kolotkin et al. 2017, 2019) is a 20-item, obesity-specific, patient-reported outcome instrument developed for use in weight management clinical studies. The IWQOL-Lite-CT assesses 2 primary domains of obesity-and health-related quality of life: Physical (7 items) and Psychosocial (13 items), and a 5-item subset of the Physical domain, the Physical Function composite. Items in the Physical Function composite describe physical impacts related to general and specific physical activities. All items are rated on either a 5-point frequency ("never" to "always") scale or a 5-point truth ("not at all true" to "completely true") scale. The 2 domain scores, 1 composite score and total score range from 0 to 100 with higher scores indicating greater functioning, derived as below: Raw scores for each subscale are computed if a minimum of 50% of the items for that subscale are nonmissing and for the IWQOL-Lite-CT total score if a minimum of 75% of all items are nonmissing. If the minimum number of items is answered for a composite, then • compute the average of the valid	Physical domain score and psychosocial domain score	Physical domain includes Items 1-5, 16, 17; Psychosocial domain includes Items 6-8, 9-15, 18, 19, 20	The minimum requirements for nonmissing responses: Psychosocial: 7 of 13 items, Physical: 4 of 7 items, otherwise considered missing
		Change from baseline in physical domain score and psychosocial domain score	Calculated as Postbaseline score - baseline score	Missing if baseline or postbaseline value is missing
		Physical function composite score	Include Items 1-3, 16, 17.	The minimum requirements for nonmissing responses: 3 of 5 items, otherwise considered missing
		Change from baseline in physical function composite score	Calculated as postbaseline score − baseline score	Missing if baseline or postbaseline value is missing
		Total score	Items 1-20	The minimum requirements for nonmissing responses: 15 of 20 items, otherwise considered missing

Approved on 26 Jun 2025 GMT

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
	nonmissing responses corresponding to the items in the total or each domain/composite will be calculated (1 = "never" or "not at all true" and 5 = "always" or "completely true"). • the score will be then calculated by transforming the raw score to the 0 (worst) to 100 (best) metric using the following formula for every participant at each time point: $100(S_{max} - C_{avg})/(S_{max} - S_{min})$ ○ C_{avg} is the raw average score of all nonmissing item responses in the composite; this average must be a number between 1 and 5, inclusive. ○ S_{max} is the maximum possible raw score value (that is, 5). ○ S_{min} is the minimum possible raw score value (that is, 1). ○ Inserting the maximum and minimum possible score values, the formula is reduced to: $100(5 - C_{avg})/4$.	Change from baseline in total score	Calculated as Postbaseline score − baseline score	Missing if baseline or postbaseline value is missing
EQ-5D-5L and CCI	The EQ-5D-5L (EuroQol Research Foundation 2019) is a standardized, 5-item, self-administered instrument for use as a measure of health outcome. It provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care as well as population health surveys. The EQ-5D-5L assesses 5 dimensions of health: Item 1: mobility Item 2: self-care	EQ-5D-5L item scores	Five health profile dimensions, each dimension has 5 levels: 1 = no problems 2 = slight problems 3 = moderate problems 4 = severe problems 5 = extreme problems It should be noted that the Numerals 1-5 have no arithmetic properties and should not be used as a primary score.	Each dimension is a single item, missing if missing. Note: Missing value can be coded as '9'.
		EQ-5D-5L UK population-based	EQ-5D-5L health states are converted into a single index 'utility' score using a scoring	If any of the items is missing or coded

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
	Item 3: usual activities Item 4: pain/discomfort Item 5: anxiety/depression Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. CCI [redacted]	utility index	algorithm based on public preferences. Uses the concatenation of the value of each EQ-5D-5L dimension score in the order: item 1; item 2; item 3; item 4; item 5. Derive EQ-5D-5L UK population-based index score according to the link[b] by using the UK algorithm (Devlin et al. 2018) to produce a patient-level index score between −0.59 and 1.0 (continuous variable)	as 9, the index score is missing.
		Change from baseline in EQ-5D-5L utility index	Calculated as postbaseline utility index score − baseline utility index score	Missing if baseline or postbaseline value is missing
		EQ-5D CCI [redacted]	Range from 0 = "worst imaginable health state" to 100 = "best imaginable health state" Note: Higher value indicates better health state.	Single item, missing if missing.
		Change from baseline in EQ-5D-5L CCI [redacted]	Calculated as CCI [redacted]	Missing if baseline or postbaseline value is missing

Abbreviations: BMI = body mass index; CKD-EPI = chronic kidney disease – epidemiology collaboration; CRF = case report form; CSV = comma separated values; DBP = diastolic blood pressure; EQ-5D-5L = EuroQol 5 Dimensions 5 levels; eGFR = estimated glomerular filtration rate; FFA = free fatty acid; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; HOMA = homeostasis model assessment; hsCRP = high-sensitivity C-reactive protein; IWQOL-Lite-CT = impact of weight on quality of life-lite clinical trials; LDL = low-density lipoprotein; MCS = Mental Component Score; PCS = Physical Component Score; PGI-C= patient global impression of change; PGI-S = patient global impression of severity; SBP = systolic blood pressure; SF-36 = Short Form 36-item Health Survey; SMBG = self-monitored blood glucose; UACR = urine albumin-creatinine ratio; CCI [redacted] VLDL = very low-density lipoprotein.

[a] Derive HOMA2 B and HOMA2 IR using the calculator at https://www.rdm.ox.ac.uk/about/our-clinical-facilities-and-mrc-units/DTU/software/homa.

[b] Derive EQ-5D-5L UK population-based index score using the calculator at https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/crosswalk-index-value-calculator/.

Table GZGQ.4.6. Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Estimand Data Points Set/Population[a]	Analysis Method	Time Point[b]	Analysis Type
Body weight	Percent change from baseline in body weight (%)	Treatment Regimen	ANCOVA with PMI	Week 72 visit	Primary endpoint, main analysis
		Efficacy	MMRM	Week 72 visit	Primary endpoint, main analysis
		Treatment Regimen	ANCOVA with MI-TP	Week 72 visit	Primary endpoint, sensitivity analysis
		Treatment Regimen	CDF plot with PMI	Week 72 visit	Exploratory analysis
		Efficacy	CDF plot with MI	Week 72 visit	Exploratory analysis
		Treatment Regimen	Waterfall Plot with observed data	Week 72 visit	Exploratory analysis
	Percentage of participants with ≥x% body weight reduction from baseline, where x = 5, 10, 15	Treatment Regimen	Logistic regression with PMI	Week 72 visit	Key secondary, main analysis
		Efficacy	Logistic regression with MI	Week 72 visit	Key secondary, main analysis
	Percentage of participants with ≥20% body weight reduction from baseline	Treatment Regimen	Logistic regression with PMI	Week 72 visit	Exploratory analysis
		Efficacy	Logistic regression with MI	Week 72 visit	Exploratory analysis
	Change from baseline in body weight (kg)	Efficacy	MMRM	Week 72 visit	Additional secondary
		Treatment Regimen	ANCOVA with PMI	Week 72 visit	Additional secondary
BMI	Change from baseline in BMI (kg/m^2)	Efficacy	MMRM	Week 72 visit	Additional secondary
		Treatment Regimen	ANCOVA with PMI	Week 72 visit	Additional secondary
	Percentage of participants with BMI threshold value of $<x\ kg/m^2$, where x = 25, 27, 30, 35	Efficacy	Logistic regression with MI	Week 72 visit	Exploratory analysis
		Efficacy	Logistic regression with PMI	Week 72 visit	Exploratory analysis
Waist circumference	Change from baseline in waist circumference (cm)	Treatment Regimen	ANCOVA with PMI	Week 72 visit	Key secondary, main analysis
		Efficacy	MMRM	Week 72 visit	Key secondary, main analysis

Approved on 26 Jun 2025 GMT

Blood pressure	Change from baseline in SBP (mmHg) Change from baseline in DBP (mmHg)	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Key secondary, main analysis; exploratory analysis
		Efficacy	MMRM (pooled and individual OFG doses)	Week 72 visit	Key secondary, main analysis; exploratory analysis
Lipid parameter	Percent (%) change from baseline in • triglycerides • non−HDL-cholesterol • total cholesterol • LDL-cholesterol • HDL-cholesterol • VLDL-cholesterol	Treatment Regimen	ANCOVA with PMI[c] (log transformation; pooled and individual OFG doses)	Week 72 visit	Key secondary, main analysis; exploratory analysis
		Efficacy	MMRM (log transformation; pooled and individual OFG doses)	Week 72 visit	Key secondary, main analysis; exploratory analysis
Glycemic control	Change from baseline in fasting glucose (mg/dL and mmol/L)	Treatment Regimen	ANCOVA with PMI	Week 72 visit	Key secondary, main analysis
		Efficacy	MMRM	Week 72 visit	Key secondary, main analysis
	Change from baseline in HbA1c (% and mmol/mol)	Treatment Regimen	ANCOVA with PMI	Week 72 visit	Key secondary, main analysis
		Efficacy	MMRM	Week 72 visit	Key secondary, main analysis
	Percentage of participants with HbA1c target value of ≤6.5%	Treatment Regimen	Logistic regression with PMI	Week 72 visit	Key secondary, main analysis
		Efficacy	Logistic regression with MI	Week 72 visit	Key secondary, main analysis
	Percentage of participants with HbA1c target value of <7%	Treatment Regimen	Logistic regression with PMI	Week 72 visit	Key secondary, main analysis
		Efficacy	Logistic regression with MI	Week 72 visit	Key secondary, main analysis
	Percentage of participants with HbA1c target value of <5.7%	Treatment Regimen	Logistic regression with PMI	Week 72 visit	Additional secondary
		Efficacy	Logistic regression with MI	Week 72 visit	Additional secondary
	Change from baseline in mean of all 7-point SMBG (mg/dL)	Efficacy	MMRM	Week 72 visit	Exploratory analysis

Approved on 26 Jun 2025 GMT

	Change from baseline in point x SMBG (mg/dL), where x = 1, …, 7	Efficacy	Descriptive statistics	Week 72 visit	Exploratory analysis
	Change from baseline in mean of premeal SMBG (mg/dL)	Efficacy	MMRM	Week 72 visit	Exploratory analysis
	Change from baseline in postmeal SMBG (mg/dL)	Efficacy	MMRM	Week 72 visit	Exploratory analysis
Fasting Insulin	Percent change baseline in fasting insulin (%)	Efficacy	MMRM (log transformation)	Week 72 visit	Additional secondary
		Treatment Regimen	ANCOVA with PMI (log transformation)	Week 72 visit	Additional secondary
Renal function	Percent change from baseline in UACR (%)	Efficacy	MMRM (log transformation)	Week 72 visit	Exploratory analysis
	Change from baseline in eGFR CKD-EPI Cystatin-C (ml/min/1.73m^2)	Efficacy	MMRM	Week 72 visit	Exploratory analysis
	Change from baseline in eGFR CKD-EPI Creatinine (ml/min/1.73m^2)	Efficacy	MMRM	Week 72 visit	Exploratory analysis
Inflammatory biomarker	Percent change from baseline in hsCRP (%)	Efficacy	MMRM (log-transformation)	Week 72 visit	Exploratory analysis
		Treatment Regimen	ANCOVA with PMI (log transformation)	Week 72 visit	Exploratory analysis
FFAs	Percent change from baseline in FFAs (%)	Efficacy	MMRM (log-transformation)	Week 72 visit	Exploratory analysis
HOMA	Percent change from baseline in • HOMA2 B (insulin) • HOMA2 IR (insulin) • HOMA2 B (C-peptide) • HOMA2 IR (C-peptide)	Efficacy	MMRM (log transformation)	Week 72 visit	Exploratory analysis

Approved on 26 Jun 2025 GMT

SF-36	Change from baseline in Short Form 36 Version 2 Health Survey (SF-36v2) acute form Physical Functioning domain score	Treatment Regimen / Randomized participants with limitations in physical function at baseline[d]	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with limitations in physical function at baseline[d]	MMRM (pooled and individual OFG doses)	Week 72 visit	Exploratory analysis
	Change from baseline in SF-36v2 acute form domain scores, PCS and MCS	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis
		Efficacy	MMRM (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis
PGI-S	PGI-S rating	Treatment Regimen	Shift from baseline (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis
		Efficacy	Shift from baseline (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis
PGI-C	PGI-C rating	Treatment Regimen	Descriptive statistics (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis
		Efficacy	Descriptive statistics (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis
IWQOL-Lite-CT	Change from baseline in: • physical function composite score • physical domain score • psychosocial domain score • total score	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis
		Efficacy	MMRM (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis
EQ-5D-5L	Change from baseline in EQ-5D-5L utility index and CCI	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis
		Efficacy	MMRM (pooled and individual OFG doses)	Week 72 visit	Additional secondary; exploratory analysis

Approved on 26 Jun 2025 GMT

Abbreviations: ANCOVA = analysis of covariance; BMI = body mass index; CDF = cumulative distribution function; CKD-EPI = chronic kidney disease - epidemiology; DBP = diastolic blood pressure; eGFR = estimated glomerular filtration rate; EQ-5D-5L = EuroQol 5 Dimensions 5 levels; FFA = free fatty acid; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; HOMA = homeostasis model assessment; hsCRP = high-sensitivity C-reactive protein; IWQOL-Lite-CT = impact of weight on quality of life-lite clinical trials; LDL = low-density lipoprotein; MCS = Mental Component Score; MI = multiple imputation; MI-TP = multiple imputation–based tipping-point; mMI = modified multiple imputation; MMRM = mixed model for repeated measures; OFG = orforglipron; PCS = Physical Component Score; PGI-C = patient global impression of change; PGI-S = patient global impression of severity; PMI = primary multiple imputation; SBP = systolic blood pressure; SF-36 = Short Form 36-item Health Survey; SMBG = self-monitored blood glucose; UACR = urine albumin-creatinine ratio; CCI [redacted] VLDL = very low-density lipoprotein.

[a] Population is displayed if different from Randomized Participants.

[b] Assessments collected at multiple postbaseline visits will be analyzed at those scheduled visits using MMRM as supplemental analyses, in addition to the primary time point listed in the table.

[c] PMI strategy will be performed after the log transformation.

[d] The limitation in physical function at baseline is defined as PGI-S response at baseline of “moderately limited,” “very much limited,” or “extremely limited.”

4.4.1. Key Secondary Endpoints

4.4.1.1. Definition of Endpoints

The definitions of key secondary endpoints are specified in Table GZGQ.4.5.

4.4.1.2. Main Analytical Approach

The analytical approaches for the key secondary endpoints are specified in Table GZGQ.4.6.

4.4.1.3. Sensitivity Analyses

No sensitivity analyses are planned for key secondary endpoints.

4.4.1.4. Supplementary Analyses

No supplemental analyses are planned for key secondary endpoints.

4.4.2. Additional Secondary Endpoints

The definitions of additional secondary endpoints are specified in Table GZGQ.4.5. Additionally, the analytical approaches are specified in Table GZGQ.4.6.

4.5. Exploratory Endpoints/Estimands Analysis

Exploratory endpoints and analyses are specified in Table GZGQ.4.5 and Table GZGQ.4.6. Additional exploratory analyses will be described in the Exploratory/Supplemental Analyses Plan and Health Technology Analyses Plan.

4.6. Safety Analyses

The planned safety analyses are consistent with compound level safety standards, which are based on various sources, including company standards, internal and external subject matter experts, publications from cross-industry initiatives (for example, PHUSE 2013, 2015, 2017, 2018, 2022), and publications from regulatory agencies (for example, EMA 2014; CDER/BIRSS 2022a, 2022b). Descriptions of the safety analyses are provided in this SAP, but some details are found in compound level safety standards.

For safety interpretation, p-values will not be used for hypothesis testing. P-values will be considered as a number between 0 and 1 that gives an idea of how strong the evidence is for an imbalance between study intervention arms. The evidence for an imbalance is stronger toward 0 and weaker toward 1. Similarly, CIs will not be used for hypothesis testing. They reflect the uncertainty of an estimate.

Unless otherwise specified, safety analyses will be conducted using the safety participants (Table GZGQ.3.1) and the safety data points set (Table GZGQ.3.2). It is noted that for some safety endpoints, such as treatment discontinuation due to AE, and so on, the analyses will be based on data observed during treatment period only.

Summary tables with risk difference will be sorted by decreasing risk difference.

4.6.1. Extent of Exposure

Duration of exposure to study intervention will be summarized by treatment group for safety participants. Descriptive statistics for participant weeks (or days) of exposure and total participant years in exposure will be provided. Overall exposure will be summarized in total patient-year (PY) of exposure, derived in the following manner:

total PY of exposure = sum of duration of exposure in days (for all participants in treatment group) / 365.25

The frequency of participants falling into different exposure ranges will be summarized:

- >0; ≥4 weeks; ≥8 weeks; ≥12 weeks; ≥16 weeks; ≥20 weeks; ≥24 weeks; ≥36 weeks; ≥48 weeks, ≥60 weeks; and ≥72 weeks

No p-values will be reported.

4.6.2. Adverse Events

The planned summaries for AEs are provided in Table GZGQ.4.7 and are described more fully in compound level safety standards.

Table GZGQ.4.7. Summary Tables Related to Adverse Events

Analysis	Method	Population/Period
Overview of AEs, including • TEAE • SAE • death, and • permanent discontinuation from study intervention due to an AE.	Fisher's exact	Safety/TP + FP (TP for study intervention discontinuation due to an AE)
TEAEs by PT within SOC	Fisher's exact	Safety/TP + FP
TEAEs by PT	Fisher's exact	Safety/TP + FP
Maximum Severity TEAEs		Safety/TP + FP
TEAEs with incidence ≥2% by PT	Fisher's exact	Safety/TP + FP
SAEs by PT within SOC	Fisher's exact	Safety/TP + FP
Primary AEs leading to permanent discontinuation of study intervention by PT within SOC	Fisher's exact	Safety/TP
Primary AEs leading to permanent discontinuation of study by PT within SOC	Fisher's exact	Safety/TP + FP
AEs leading to study intervention interruption by PT within SOC	Fisher's exact	Safety/TP
AEs leading to study intervention reduction by PT within SOC	Fisher's exact	Safety/TP
Listing of SAEs		Safety/TP + FP
Listing of primary AEs leading to permanent discontinuation of study intervention		Safety/TP
Listing of primary AEs leading to permanent discontinuation of study		Safety/TP + FP
Listing of deaths		Safety/TP + FP
Listing of AEs related to suspected overdosing		Safety/TP

Approved on 26 Jun 2025 GMT

Analysis	Method	Population/Period
Listing of participants with at least 1 notable event		Randomized/TP + FP

Abbreviations: AE = adverse event; eCRF = electronic case report form; FP = follow-up period; PT = preferred term; SAE = serious adverse event; SOC = system organ class; TEAE = treatment-emergent adverse event; TP = treatment period.

4.6.3. Clinical Laboratory Evaluation

The planned summaries for clinical laboratory evaluations are provided in Table GZGQ.4.8 and are described more fully in compound level safety standards.

Table GZGQ.4.8. Summary Tables Related to Clinical Laboratory Evaluations

Analysis	Method	Population/Period
Box plots and mean/SD (or 95% CI) for observed values by visit	Descriptive statistics	Safety/TP + FP
Box plots and mean/SD (or 95% CI) for change from baseline values by visit	Descriptive statistics	Safety/TP + FP
Summary for participants with elevated or low values meeting specified levels	Descriptive statistics	Safety/TP + FP
Listing of abnormal laboratory findings		Safety/TP + FP

Abbreviations: CI = confidence interval; FP = follow-up period; SD = standard deviation; TP = treatment period.

4.6.4. Vital Signs and Physical Characteristics

Triplicate vital signs will be collected at the same visit, thus the mean of these measurements will be used for the vital signs analyses. The planned summaries for vital signs (systolic BP, diastolic BP, pulse rate) are provided in Table GZGQ.4.9 and are described more fully in compound level safety standards.

Table GZGQ.4.9. Summary Tables Related to Vital Signs

Analysis	Method	Population/Period
Box plots and mean/SD (or 95% CI) for observed values by visit	Descriptive statistics	Safety/TP + FP
Box plots and mean/SD (or 95% CI) for change from baseline values by visit	Descriptive statistics	Safety/TP + FP
Analysis of blood pressure and pulse rate for change from baseline	MMRM	Safety/TP + FP
Summary for participants meeting specific blood pressure and pulse rate levels	Descriptive statistics	Safety/TP + FP
Shift of maximum-to-maximum for pulse rate	Descriptive statistics	Safety/TP + FP

Abbreviations: CI = confidence interval; FP = follow-up period; MMRM = mixed model for repeated measures; SD = standard deviation; TP = treatment period.

4.6.5. Electrocardiograms

Triplicate electrocardiograms (ECGs) will be collected at the same visit; thus the mean of these measurements will be used for the analysis. The planned summaries for ECG parameters are provided in Table GZGQ.4.10 and are described more fully in compound level safety standards.

Table GZGQ.4.10. Summary Tables Related to ECG Parameters

Analysis	Method	Population/Period
Box plot and mean/SD (or 95% CI) for observed values by visit	Descriptive statistics	Safety/TP + FP
Box plot and mean/SD (or 95% CI) for change from baseline values by visit	Descriptive statistics	Safety/TP + FP
Heart rate and PR interval in specified categories	Descriptive statistics	Safety/TP + FP
Analysis of change from baseline in ECG parameters (heart rate, PR interval)	MMRM	Safety/TP + FP

Abbreviations: CI = confidence interval; ECG = electrocardiogram; FP = follow-up period; MMRM = mixed model for repeated measures; PR = pulse rate; SD = standard deviation; TP = treatment period.

4.6.6. Safety Topics of Interest

This section includes safety topics of interest whether due to observed safety findings, potential findings based on drug class, or safety topics anticipated to be requested by a regulatory agency for any reason. In general, safety topics of interest will be identified by 1 or more standardized Medical Dictionary for Regulatory Activities (MedDRA) query(ies) (SMQs), system organ class (SOC), high-level term (HLT), FDA Medical Query (FMQ), or a Lilly defined MedDRA preferred term (PT) listing based upon the review of the most current MedDRA Version, or by relevant laboratory changes. Searching criteria are detailed in the compound level safety standards.

The planned analyses for safety topics of interest are provided in Table GZGQ.4.11 and are described more fully in compound level safety standards.

Table GZGQ.4.11. Description and Analyses of Safety Topics of Interest

Special Safety Topic	Short Description	Analysis	Method	Population/Period
Major Adverse Cardiovascular Events	Death and nonfatal CV AEs will be adjudicated by a committee of physicians external to Lilly with cardiology expertise: CEC.	Positively adjudicated MACE by category/subcategory MACE reported by investigators may be summarized		Safety/TP + FP
		Listing of all MACE reported by investigator (whether or not positively adjudicated) (if necessary)		Safety/TP + FP
Arrhythmias and Cardiac Conduction Disorders	The TE arrhythmias and cardiac conduction disorders events will be derived using the MedDRA PTs contained in certain SMQs and HLT.	TE arrhythmias and cardiac conduction disorders by PT nested within SMQ and HLT[a]	Fisher's exact	Safety/TP + FP
Hypotension, Orthostatic Hypotension, and Syncope	The TE hypotension, orthostatic hypotension, and syncope events will be derived using MedDRA PTs and FMQ	TE hypotension, orthostatic hypotension, and syncope by PT[a]	Fisher's exact	Safety/TP + FP
		Hypotension by baseline antihypertensive medication use using hypotension narrow FMQ		

Hypoglycemia[b]	The American Diabetes Association position statement on glycemic targets (ADA 2025) will be used to define: • Level 1 hypoglycemia • Level 2 hypoglycemia • Level 3 hypoglycemia (severe/serious hypoglycemia) • Nocturnal hypoglycemia events (including severe hypoglycemia) **occur at night** and presumably during sleep.	Incidence (percent of patients experiencing 1 or more episode) of Level 2 or Level 3 hypoglycemia	Logistic regression for the overall postbaseline period	Safety/TP + FP excluding events after initiation of new antihyperglycemic therapy(regardless of how many days the medication was taken), as defined in Section 6.3 and Table GZGQ.6.4 Supportive analysis: Safety/TP + FP regardless of initiation of new antihyperglycemic therapy
		Incidence of Level 3 (severe/serious) hypoglycemia		
		Incidence of Level 1 hypoglycemia (if warranted by data)[a]		
		Incidence of Level 2 or Level 3 nocturnal hypoglycemia		
		Rate (episodes/patient/year) of Level 2 or Level 3 hypoglycemia	Negative binomial regression or empirical method for the overall postbaseline period	
		Rate of Level 3 (severe/serious) hypoglycemia		
		Rate of Level 1 hypoglycemia (if warranted by data)		
		Rate of Level 2 or Level 3 nocturnal hypoglycemia		
		Listing of Level 2 or Level 3 hypoglycemia events		
		Listing of Level 3 or Level 3 nocturnal hypoglycemia events		
Severe Persistent Hyperglycemia	Initiation of new antihyperglycemic therapy taken for severe persistent hyperglycemia	Rescue therapy taken for severe persistent hyperglycemia (if warranted by data)[a]	Fisher's exact	Safety/TP + FP
Gastrointestinal (GI) Safety[d]	GI AEs using Gastrointestinal disorders SOC and FMQ will be captured.	Severe or serious TE GI events by PT	Fisher's exact	Safety/TP + FP

		Study intervention discontinuation due to TE GI events	Fisher's exact	Safety/TP + FP
		TE abdominal pain by FMQ narrow PT	Fisher's exact	Safety/TP + FP
		Prevalence and incidence over time for TE nausea, vomiting, diarrhoea, constipation, and NVD		Safety/TP + FP
		Plot of time to the onset of TE nausea, vomiting, diarrhoea, constipation, and NVD	KM	Safety/TP + FP
		Plot of prevalence and incidence over time for TE nausea, vomiting, diarrhoea, constipation, and NVD by maximum severity		Safety/TP + FP
Renal Safety	Laboratory measures related to renal safety will be analyzed. Renal events, including acute renal failure and chronic renal failure exacerbation, will be captured using SMQs. Dehydration events will be captured using SMQ.	Shift of minimum-to-minimum for eGFR estimated by the CKD-EPI Cystatin C equation[c]		Safety/TP + FP
		Shift of maximum-to-maximum for UACR		Safety/TP + FP
		Summary of eGFR decrease from baseline		Safety/ TP + FP
		MMRM analyses for eGFR estimated by the CKD-EPI Cystatin C equation[c]	MMRM	Safety/TP + FP
		MMRM analyses for UACR (log transformation)	MMRM	Safety/TP + FP
		TE renal events by PT nested within SMQ[a]	Fisher's exact	Safety/TP + FP
		Listing of TE dehydration events by PT[a]		Safety/TP + FP
Pancreatic Safety	The pancreatic enzyme data (p-amylase and lipase) will be observed through laboratory testing. All suspected cases of acute or chronic pancreatitis and AEs of severe or serious	Shift of maximum-to-maximum for pancreatic enzyme		Safety/TP + FP
		MMRM analysis for pancreatic enzymes (p-amylase and lipase) with a log transformation (postbaseline/baseline)	MMRM	Safety/TP + FP

	abdominal pain of unknown etiology will be sent for adjudication by an independent clinical endpoint committee.			
		Investigator-reported events and subsequently confirmed adjudicated events, respectively	Fisher's exact	Safety/TP + FP
		TE Pancreatitis events by "Acute pancreatitis" SMQ (narrow scope) and "Chronic pancreatitis" PT		Safety/TP + FP
		Listing of adjudicated and investigator-reported pancreatic events		Safety/TP + FP
Thyroid Safety	TE thyroid malignancies and C-cell hyperplasia will be identified using MedDRA HLT and PT. The purpose of calcitonin measurements is to assess the potential effect of orforglipron on thyroid C-cell function.	TE thyroid C-cell hyperplasia and malignancies by PT within Event Category	Fisher's exact	Safety/TP + FP
		Shift of maximum-to-maximum for calcitonin value in the thresholds		Safety/TP + FP
		Listing of participants who meet protocol defined discontinuation criteria based on calcitonin: • with eGFR <60 mL/min/1.73 m^2 at baseline, serum calcitonin value ≥35 ng/L, AND ≥50% increase from the baseline value[a] • with eGFR ≥60 mL/min/1.73 m^2 at baseline, serum calcitonin value ≥20, and <35 ng/L AND ≥50% increase from the baseline value and ≥10% increase on retest[a] • with eGFR ≥60 mL/min/1.73 m^2 at baseline, serum calcitonin value ≥35 ng/L, AND ≥50% increase from the baseline value[a]		Safety/TP + FP Safety/TP + FP Safety/TP + FP
		MMRM analysis for calcitonin (log transformation)	MMRM	Safety/TP + FP
Malignancies	The malignancy events will be derived using the MedDRA PTs contained in certain SMQs.	TE malignancy by PT nested within SMQ[a]	Fisher's exact	Safety/TP + FP
Hepatic Safety	Hepatic labs include ALT, AST, ALP, TBL, DBL, and GGT. When criteria are met for hepatic evaluations, investigators will conduct close monitoring of hepatic symptoms and liver	Abnormal postbaseline categories for hepatic safety parameters: ALT, AST, ALP, TBL, DBL, and GGT		Safety/TP + FP
		TE potentially drug-related hepatic disorders by PT nested within SMQs	Fisher's exact	Safety/TP + FP
		Hepatocellular drug-induced liver injury screening plot (maximum TBL vs maximum ALT or AST)		Safety/TP + FP

	tests, perform a comprehensive evaluation for alternative causes of abnormal liver tests, and complete follow-up hepatic safety eCRFs. Potentially drug-related hepatic disorders will be identified using SMQs.	Hepatocellular drug-induced liver injury screening table		Safety/TP + FP
		Cholestatic drug-induced liver injury screening table		Safety/TP + FP
		Cholestatic drug-induced liver injury screening plot (maximum TBL vs maximum ALP)		Safety/TP + FP
		Hepatic Laboratory Parameters (ALT, AST, ALP, TBL, DBL, and GGT) with a log transformation (postbaseline/baseline)	MMRM	Safety/TP + FP
		Plot of Hepatic Laboratory Parameters (ALT, AST, ALP, TBL, DBL, and GGT)		Safety/ TP + FP
		Listing of participants with ALT or AST ≥3X ULN		Safety/TP + FP
		Listing of participants with ALP or TBL ≥2X ULN		Safety/TP + FP
		Shift of maximum-to-maximum for ALT, AST, ALP, TBL		Safety/TP + FP
		Participant profiles for participants meeting criteria for a comprehensive hepatic evaluation (as defined in the protocol)		Safety/TP + FP
Gallbladder and Biliary Tract Disorders	All events of TE gallbladder and biliary tract disorders will be identified by using certain SMQs.	TE gallbladder and biliary tract disorders by PT nested within SMQ[a]	Fisher's exact	Safety/TP + FP
Hypersensitivity Reactions	All events of TE allergic reaction and hypersensitivities will be identified by using certain SMQs[a]	TE allergic reaction and hypersensitivities by PT nested within applicable SMQ[a]	Fisher's exact	Safety/TP + FP
Depression, Suicidal Ideation, and Behavior	AEs will be searched using MedDRA PTs from certain SMQs or FMQ[a].	TE major depressive disorder, suicidal ideation, or behavior events by PT nested within applicable SMQ[a] or FMQ[a]	Fisher's exact	Safety/TP + FP
	Suicide-related thoughts and behaviors will be collected based on the C-SSRS.	Summary of C-SSRS categories and composite measures		Safety/TP + FP
		Listing of C-SSRS categories and composite measures		Safety/TP + FP

	Patient health questionnaire-9 (PHQ-9) will be collected to assess the specific diagnostic symptoms that determine the presence of a clinical depressive disorder. The PHQ-9 total scores will be categorized as minimal to none, mild, moderate, moderately severe, and severe.	Shift of each baseline category (maximum value) versus each postbaseline category (maximum value)		Safety/TP + FP
		Summary of categories based on the maximum values during baseline and postbaseline: • any increase in depression category (that is, worsening of depression) • increase from minimal to none or mild depression to moderate, moderately severe, or severe depression • increase from mild or moderate depression to moderately severe or severe depression		Safety/TP + FP
Abuse Potential	AEs will be searched using a modified study agent abuse potential FMQ.	TE abuse potential events by PT[a]	Fisher's exact	Safety/TP + FP
AEs possibly related to loss of lean muscle mass	AEs will be identified using MedDRA HLTs and PTs	TEAEs possibly related to loss of lean muscle mass by PT within Event Category	Fisher's exact	Safety/TP + FP
Malnutrition	AEs will be identified using MedDRA HLTs and PTs	TE malnutrition by PT within Event Category	Fisher's exact	Safety/ TP +FP
Excessive weight loss	AEs will be identified using MedDRA PTs BMI< 18.5 kg/m^2 will be analyzed	Severe or serious TE excessive weight loss by PT or BMI <18.5kg/m^2	Fisher's exact	Safety/ TP + FP
		Listing of Participants with BMI< 18.5 kg/m^2		Safety/ TP + FP
Diabetic Retinopathy	Diabetic retinopathy and macular edema will be monitored using retinal fundus photography and additional ophthalmological follow-up as warranted throughout the study. Diabetic retinopathy, macular edema, and related complication will be searched using MedDRA PTs.	Baseline and postbaseline fundus photography exam by treatment group and by eye (that is, any eye, left eye, and right eye)		Safety/TP + FP
		Shift of baseline diabetic retinopathy and macular edema status to most severe postbaseline diabetic retinopathy and macular edema status		Safety/TP + FP
		TE diabetic retinopathy complications by PT	Fisher's exact	Safety/TP + FP

Approved on 26 Jun 2025 GMT

Abbreviations: AE = adverse event; ALP = serum alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; CEC = clinical endpoint committee; CKD-EPI = Chronic Kidney Disease Epidemiology; C-SSRS = Columbia Suicide Severity Rating Scale; CV = cardiovascular; DBL = direct bilirubin; eCRF = electronic case report form; eGFR = estimated glomerular filtration rate; FMQ = FDA Medical Query; FP = follow-up period; GGT = gamma-glutamyl transferase; HLT = high-level term; KM = Kaplan-Meier; MACE = major adverse cardiovascular event; max-to-max = maximum to maximum; MedDRA = Medical Dictionary for Regulatory Activities; min-to-min = minimum to minimum; MMRM = mixed model for repeated measures; NVD = ;nausea, vomiting, diarrhea; PT = preferred term; SMQ = standardized MedDRA query; SOC = system organ class; TE = treatment-emergent; TBL = total bilirubin; TP = treatment period; UACR = urinary albumin-to-creatinine ratio.

[a] For these tables, if the number of events is low, a listing may be provided instead.

[b] Glucose values collected on the same date as an OGTT assessment are not included

[c] eGFR will be calculated by Chronic Kidney Disease-Epidemiology cystatin-c equation (Inker et al, 2012).

[d] Additionally, all GI events will be analyzed.

Note: Listings and participant profiles may be provided through interactive display tools instead of a static display.

4.7. Other Analyses

4.7.1. Subgroup Analyses

Subgroup analyses will be conducted for the primary endpoint with the treatment regimen estimand:

- percent change in body weight from baseline at Week 72 visit.

The variables for subgroup analysis are specified in Section 6.1 and Table GZGQ.6.1.

The ANCOVA model specified in Section 4.1.2.1.1, will be fitted separately within each subgroup. The LS mean, LS mean difference, standard error, and 95% CI will be presented. The LS means and variance-covariance estimates from these separate models will be used to test the treatment-by-subgroup interaction at the significance level of 0.10.

If the subgroup factor is part of the strata variable, a new strata variable that excludes the subgroup categories will be used.

If any category within the subgroup is <5% of the total population, only descriptive statistics will be provided for that category (that is, no inferential testing within the subgroup category).

A forest plot including the treatment difference and 95% CI estimated for each subgroup level will be presented.

A Bayesian shrinkage method will be used to provide adjusted estimates and improve estimation precision. Following the paper by Wang et.al. (2024), for a given baseline characteristic with k subgroups, let Y_i $(i = 1, \ldots, k)$ be the observed sample estimate of the treatment effect difference in subgroup i. The following hierarchical model will be used:

$$Y_i \sim N(\mu_i, \sigma_i^2),$$

$$\mu_i \sim N(\mu, \tau^2),$$

$$\mu \sim N(0, 30^2),$$

$$\tau \sim N^+(10),$$

where σ_i^2 is set to the observed variance for sample estimate, $N^+(10)$ is the half-normal distribution with a scaled parameter of 10. A weakly informative prior, as suggested by Röver et al. (2021) and specified above, will be applied to μ and τ. Of note, the unit information standard deviation (UISD) is assumed to be 15% based on historical weight management studies, and a standard deviation of 30% is chosen for the centrality parameter μ , so that it is approximately twice the UISD. A scale parameter of 10 is chosen for the half- normal distribution such that the median heterogeneity in percent body weight reduction is approximately 6.7%, which is a reasonably large difference.

Additional subgroup evaluations may be conducted as exploratory analyses.

4.8. Interim Analyses

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

4.8.1. Data Monitoring Committee

A data monitoring committee (DMC) is not planned for this study.

An independent DMC will be established for interim safety monitoring of Study J2A-MC-GZGS (ACHIEVE-4) (GZGS), "A Phase 3, Open Label, Study of Once-Daily LY3502970 Compared with Insulin Glargine in Adult Participants with Type 2 Diabetes and Obesity or Overweight at Increased Cardiovascular Risk", a study within the orforglipron Phase 3 program for chronic weight management (CWM) and T2D. The DMC will have the responsibility to review unblinded interim safety analysis results from that study. The DMC may be asked to review unblinded safety data from Study GZGQ if a need arises following the blinded trial-level safety reviews (TLSRs) conducted by the sponsor. If needed, permanent data transfers will occur for the purposes of supporting the DMC. Only the statisticians from the statistical analysis center (SAC) will have access to the unblinded data that are presented to the DMC. The SAC and members of the DMC will abide by the principles and responsibilities described in the Study GZGS DMC charter, which includes keeping all unblinded information confidential until the planned unblinding of the trial.

4.9. Changes to Protocol-Planned Analyses

There are changes to the analyses described in the protocol:

- orforglipron 6 mg and 12 mg endpoint "mean change in waist circumference (cm)" was removed from the key secondary objectives and added to the additional secondary objectives.
- treatment compliance is defined as taking at least 75% of required study intervention during the treatment period.

Approved on 26 Jun 2025 GMT

5. Sample Size Determination

A sample size of 1500 participants (300 participants per orforglipron treatment group and 600 placebo participants) provides more than 90% power to demonstrate superiority of 36 mg, 12 mg, and/or 6 mg orforglipron to placebo with regards to mean percent change in body weight from baseline to Week 72. The sample size determination assumes that evaluation of superiority of 36 mg, 12 mg, and 6 mg orforglipron to placebo will be conducted in parallel, each at a 2-sided significance level of 0.016 using a 2-sample t-test for the "treatment regimen" estimand. Additionally, a difference of at least 5% mean body weight reduction at 72 weeks for 36 mg, 12 mg, and 6 mg orforglipron compared with placebo, a common SD of 10%, and a dropout rate of 30% in the placebo group and 20% in the orforglipron groups are assumed for statistical power calculations.

Approved on 26 Jun 2025 GMT

6. Supporting Documentation

6.1. Appendix 1: Demographic and Baseline Characteristics

Demographics and other baseline characteristics will be summarized by treatment group for randomized participants. The listing of basic demographic characteristics (that is, age, sex, ethnicity, race, country, baseline body weight, and so on) for randomized participants will be provided.

The continuous variables will be summarized using descriptive statistics, and the categorical variables will be summarized using frequency counts and percentages.

Table GZGQ.6.1 describes the specific variables and how they will be summarized. The last column specifies variables used for the subgroup analysis described in Section 4.7.1.

Table GZGQ.6.1. Demographics and Baseline Characteristics with Variables for Subgroup Analysis

Variable	Quantitative Summary	Categorical Summary	Subgroup Analysis[a]
Demographics			
Age[b]	Yes	<65, ≥65 years	X
		<75, ≥75 years	X
		<65, ≥65 and <75, ≥75 and <85, ≥85 years	
Sex	No	Male, Female	X
Ethnicity	No	Hispanic/Latino, Non-Hispanic/Non-Latino	X
Race	No	American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian or other Pacific Islander, White, or Multiple	X
Country	No	By each country	X
Geographic region	No	North America, Europe, Australia, Asia or Central/South America	
			X
Height (cm)	Yes		
Baseline waist circumference (cm)	Yes		
Baseline body weight (kg)	Yes		
Baseline BMI	Yes	≥27 to <30, ≥30 to <35, ≥35 to <40, ≥40 kg/m^2	X
		≤ median, > median	
Caffeine use	No	Never, Current, Former	
Alcohol use	No	Never, Current, Former	
Tobacco use	No	Never, Current, Former	
Nicotine Replacement use	No	Never, Current, Former	
Recreational drug use - methamphetamine	No	Never, Current, Former	
Recreational drug use - cocaine	No	Never, Current, Former	
Baseline Therapy			
Baseline Background oral AHMs	No	Weight loss, Weight gain, Weight neutral and others	X
Baseline number of AHMs	No	0, 1, 2, 3	
Baseline Disease Characteristics			

Baseline systolic blood pressure (mmHg)	Yes		
Baseline diastolic blood pressure (mmHg)	Yes		
Baseline pulse rate (beats/min)	Yes		
Baseline eGFR calculated by CKD-EPI Cystatin-C equation (mL/min/1.73m^2)	Yes	<60, ≥60 mL/min/1.73m^2	
Baseline eGFR CKD-EPI Creatinine (mL/min/1.73m^2)[c]	Yes	<60, ≥60 mL/min/1.73m^2	
Baseline UACR (g/kg)	Yes	<30, ≥30 and ≤300, >300 g/kg	
Baseline triglycerides (mg/dL)	Yes		
Baseline total cholesterol (mg/dL)	Yes		
Baseline VLDL-cholesterol (mg/dL)	Yes		
Baseline non–HDL-cholesterol (mg/dL)	Yes		
Baseline LDL-cholesterol (mg/dL)	Yes		
Baseline HDL-cholesterol (mg/dL)	Yes		
Baseline fasting insulin (pmol/L)	Yes		
Baseline HbA1c (%)	Yes	≤8.5% (≤69 mmol/mol), >8.5% (>69 mmol/mol)	X
Baseline HbA1c (mmol/mol)	Yes		
Baseline fasting glucose (mg/dL)	Yes		
Baseline fasting glucose (mmol/L)	Yes		
Duration of diagnosis for T2D (years)	Yes	≤5, >5 and ≤10, >10 years	
Duration of obesity (years)	Yes		

Abbreviations: AHM = antihyperglycemic medication; BMI = body mass index; CKD-EPI = chronic kidney disease - epidemiology; eCRF = electronic case report form; eGFR = estimated glomerular filtration rate; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; LDL = low-density lipoprotein; OSAS = obstructive sleep apnea syndrome; T2D = type 2 diabetes; UACR = urine albumin-to-creatinine ratio; VLDL = very low-density lipoprotein.

[a] Subgroup analyses are defined in Section 4.7.1 with more details.

[b] Age in years will be calculated as length of the time interval from the imputed date of birth (1 July in the year of birth collected in the eCRF) to the informed consent date.

[c] eGFR will be calculated by Chronic Kidney Disease-Epidemiology cystatin-c equation (Inker et al, 2012).

6.2. Appendix 2: Historical Illnesses and Preexisting Conditions

Historical illness is defined as the condition/event with an end date prior to the date of informed consent.

Preexisting condition is defined as the condition/event with a start date prior to the first dose of the study intervention and stop date that is at or after the informed consent date or has no stop date (that is, ongoing). Randomization date will be used if a participant has not been dosed.

The planned summaries for historical illness and preexisting conditions are provided in Table GZGQ.6.2.

Table GZGQ.6.2. Summary Tables Related to Historical Illness and Preexisting Conditions

Analysis	Population/Period
Historical illness by PT within SOC	Randomized
Preexisting conditions by PT within SOC	Randomized
Obesity related health disorders[a]	Randomized
Comorbidities at baseline[b]	Randomized

Abbreviations: PT = preferred term; SOC = system organ class.

[a] Include the 11 Obesity-Related Health Disorders defined in Table GZGQ.6.6.

[b] CCI [redacted].

6.3. Appendix 3: Prior/Concomitant Medications

Medications that start before or at the last date of treatment period or follow-up period and are ongoing or ended during the treatment period or follow-up period will be classified as concomitant medication. Medications that start and end before first dose date of study intervention will be classified as prior therapy.

Baseline is defined as the corresponding medication taken on the day before the date of the first dose of study intervention.

If there are no doses of study intervention, randomization date will be used instead of the first dose date.

The planned summaries for concomitant medications are provided in Table GZGQ.6.3.

Table GZGQ.6.3. Summary Tables Related to Concomitant Medications

Analysis	Method	Population/Period
CMs by PN		Randomized/TP + FP
CMs by PN within ATC code		Randomized/TP + FP
Antihypertensive CMs at baseline by PN within ATC code		Randomized
Lipid lowering CMs at baseline by PN within ATC code		Randomized
Antihyperglycemic CMs at baseline by PN within ATC code		Randomized
Status change of antihypertensive and lipid lowering CMs		Randomized/TP
Antihyperglycemic CM initiated after baseline by PN within ATC code		Randomized/TP
Antidiarrheal, Antiemetic, and Constipation CMs at baseline by PN within ATC code		Randomized/TP
Antidiarrheal, Antiemetic, and Constipation CMs initiated after baseline by PN within ATC code		Randomized/TP
Prohibited concomitant medication and surgical procedure on weight management treatment by PN within ATC code		Randomized Participants/TP
Concomitant medication on glycemic rescue therapy[a] or prohibited glycemic therapy[b] by PN within ATC code		Randomized Participants/TP
Obesity management medications 1 year prior to screening by PN within ATC code		Randomized Participants

Abbreviations: ATC = Anatomical Therapeutic Chemical; CM = concomitant medication; FP = follow-up period; PN = preferred name; TP = treatment period.

[a] Glycemic rescue therapy refers to the initiation of a new glycemic therapy marked as rescue and continued for more than 2 weeks (14 days) or an increase in the dose of an existing glycemic therapy marked as rescue and continued for more than 2 weeks (14 days).

[b] The use of prohibited glycemic therapy is flagged from the initiation day.

Table GZGQ.6.4 provides the protocol-specified concomitant medications of interest.

Table GZGQ.6.4. Protocol-Specified Concomitant Medications

Category	Preferred Name/ATC Code
Antihypertensive Medication	All medications with ATC code containing 'C01', 'C03', 'C07', 'C08', or 'C09'
Lipid Lowering Medication	All medications with ATC code containing 'C10'
Antihyperglycemic Medication	All medications with ATC code containing 'A10', except for 'A10XA'

Abbreviation: ATC = Anatomical Therapeutic Chemical.

Table GZGQ.6.5 provides the protocol-specified prohibited concomitant medication and surgical procedure rules.

Table GZGQ.6.5. Prohibited Concomitant Medication Rules

Timing	Preferred Name/ATC Code	ATC Classification	Rule
Prohibited Concomitant Medication on Weight Management Treatment[a]			
Postbaseline	All medications with ATC code of A08AA	Centrally acting antiobesity products	Any initiation is prohibited during treatment period
	All medications with ATC code of A08AB	Peripherally acting antiobesity products	Any initiation is prohibited during treatment period
	All medications with ATC code of A08AX	Other antiobesity drugs	Any initiation is prohibited during treatment period
	All medications with ATC code of A08AW		Any initiation is prohibited during treatment period
	All medications that contain Phenylpropanolamine in the preferred term or Tirzepatide, Semaglutide, or Liraglutide in the preferred term		Any initiation is prohibited during treatment period
Prohibited Surgical Procedure on Weight Management Treatment[a]			
Postbaseline	ABDOMINOPLASTY		Any initiation is prohibited during treatment period
	GASTRIC BANDING		Any initiation is prohibited during treatment period
	GASTRIC BYPASS		Any initiation is prohibited during treatment period

Approved on 26 Jun 2025 GMT

Timing	Preferred Name/ATC Code	ATC Classification	Rule
	LIPOSUCTION		Any initiation is prohibited during treatment period
	SLEEVE GASTRECTOMY		Any initiation is prohibited during treatment period
	CRYOLIPOLYSIS		Any initiation is prohibited during treatment period
	MUCOSAL ABLATION		Any initiation is prohibited during treatment period
	GASTRIC ARTERY EMBOLIZATION		Any initiation is prohibited during treatment period
	INTRAGASTRIC BALLOON		Any initiation is prohibited during treatment period
	DUODENAL-JEJUNAL ENDOLUMINAL LINER		Any initiation is prohibited during treatment period
Prohibited Concomitant Medication on Glycemic Therapy[a]			
Postbaseline	All medications with ATC code of A10BH	DPP-4 inhibitors	Any initiation is prohibited during treatment period
	All medications with ATC code of A10BJ	GLP-1 analogues	Any initiation is prohibited during treatment period
	All medications with ATC code of A10BX	Other blood glucose lowering drugs, excluding insulins	Any initiation is prohibited during treatment period
	All medications that contain tirzepatide, semaglutide, or liraglutide in the preferred term		Any initiation is prohibited during treatment period

Abbreviations: ATC = Anatomical Therapeutic Chemical; DPP-4 = dipeptidyl peptidase 4; GLP-1 = glucagon-like peptide-1; ICE = intercurrent event.

[a] Use of these medications will be reviewed by the study team (blinded to study intervention) on a case-by-case basis to assess the clinical significance of the change in medication. After review, if deemed to be nonclinically significant, the case won't be considered as an ICE.

6.4. Appendix 4: Treatment Compliance

Treatment compliance is defined as taking at least 75% of required study intervention during the treatment period. Compliance will be calculated by

*([total number of doses dispensed – total number of doses returned]/total number of doses expected to be administered)*100%.*

Frequency counts and percentages of participants compliant to study intervention will be summarized by treatment group using the safety participants during treatment period. No inferential analysis will be performed.

Participants with dose interruptions/modifications will be summarized with reasons. Interruptions will be summarized if the duration of the interruption is 7 days or longer for all reasons other than AE.

6.5. Appendix 5: Important Protocol Deviations

Important protocol deviations are identified in the Trial Issues Management Plan. A listing and summary of important protocol deviations by treatment group will be provided for all randomized participants. No inferential analysis will be performed.

6.6. Appendix 6: Clinical Trial Registry Analyses

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

Analyses provided for the CTR requirements include the following:

- Summary of AEs, provided as a dataset that will be converted to a XML file. Both serious adverse events (SAEs) and 'Other' nonserious AEs are summarized by treatment group and MedDRA PT.
  - An AE is considered 'Serious' whether or not it is a treatment-emergent adverse event (TEAE).
  - An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term, and
    - the number of events experienced.
  - For each SAE, these additional terms are provided for European Union Drug Regulating Authorities Clinical Trial (EudraCT):
    - the total number of occurrences causally related to treatment
    - the total number of deaths, and
    - the total number of deaths causally related to treatment.
  - Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of participants 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: ≥18 to <65 years, ≥65 to <85 years, and ≥85 years.

6.7. Appendix 7: Statistical Analysis for China

General considerations

Analyses will be performed for the following subpopulation:

- Participants enrolled in China
- Participants enrolled in East Asian Countries/regions (China, Korea)

The analysis methods for the above-mentioned subgroups will be similar to those described for the main part of this SAP GZGQ. If there is not sufficient number of participants in the subpopulation, summary statistics will be provided.

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

6.8. Appendix 8: Statistical Analysis for Japan

The statistical analysis method for subpopulation analysis to be used for publication is described below.

Disposition, demographics, the primary endpoint, mean percent change in body weight from baseline, percentage of participants who achieve ≥5% body weight reduction, selected secondary and safety endpoints will be analyzed by the Japan Society for the Study of Obesity (JASSO) subpopulation. The analyses to be included will be documented in a list of analyses. If there is not a sufficient number of participants in this subpopulation, summary statistics will be provided. The JASSO subpopulation analysis will be performed according to the criteria of both BMI- and obesity-related health disorders according to the obesity disease label in Japan.

The target population is defined as participants who are diagnosed as having either hypertension, dyslipidemia, or T2D with:

1) baseline BMI between 27 to <35 kg/m^2, with at least 2 obesity-related health disorders, and
2) baseline BMI between ≥35 kg/m^2,

where the obesity-related health disorders are defined according to obesity disease treatment guideline (JASSO 2022).

There are 11 obesity-related health disorders, as listed below (Table GZGQ.6.6). All participants and participants with obesity disease according to the JASSO guideline will be compared.

Eleven obesity-related health disorders

The JASSO guideline defines 11 health problems for the diagnosis of “obesity disease” for participants who need weight reduction for medical reasons:

Table GZGQ.6.6 The 11 Obesity-Related Health Disorders

Obesity-Related Health Disorders	Medical History Term
CCI	

Approved on 26 Jun 2025 GMT

7. References

Andersen SW, Millen BA. On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharm Stat*. 2013;12(1):7-16. https://doi.org/10.1002/pst.1548

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

[CDER/BIRRS] Center for Drug Evaluation and Research and Biomedical Informatics and Regulatory Review Science Team. Advancing premarket safety analytics workshop meet material_FDA-DM FMQs. Published September 06, 2022a. Accessed January 24, 2024. https://www.regulations.gov/document/FDA-2022-N-1961-0001

[CDER/BIRRS] Center for Drug Evaluation and Research and Biomedical Informatics and Regulatory Review Science Team. Standard safety tables and figures: integrated guide. Published September 08, 2022b. Accessed January 24, 2024. https://www.regulations.gov/document/FDA-2022-N-1961-0046

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

Du Y, Li J, Raha S, Qu Y. A unified Bayesian framework for bias adjustment in multiple comparisons from clinical trials. *Stat Med*. 2024;43(15):2928-2943. https://doi.org/10.1002/sim.10064

Devlin N J, Shah KK, Feng Y, et al. Valuing health-related quality of life: an EQ-5D-5 L value set for England. *Health Econ*. 2018, 27(1): 7-22. https://doi.org/10.1002/hec.3564

American Diabetes Association Professional Practice Committee. 6. Glycemic Goals and Hypoglycemia: Standards of Care in Diabetes-2025. *Diabetes Care*. 2025;48(Suppl1):S128-S145. https://doi.org/10.2337/dc25-s006

[EMA] European Medicines Agency. Guidance document for the content of the <Co-> Rapporteur day 80 critical assessment report. EMA/269176/2014. Published 2014. Accessed May 3, 2023. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/day-80-assessment-report-clinical-guidance_en.pdf

EuroQol Research Foundation. EQ-5D-5L user guide, version 3.0. Updated September 2019. Accessed July 29, 2022. https://euroqol.org/publications/user-guides

[FDA] United States Food and Drug Administration. Adjusting for covariates in randomized clinical trials for drugs and biological products: guidance for industry. Published May 2023. Accessed July 24, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products

[FDA] United States Food and Drug Administration. Impact story: using innovative statistical approaches to provide the most reliable treatment outcomes information to patients and clinicians. Published June 10, 2019. Accessed December 14, 2023.

https://www.fda.gov/drugs/regulatory-science-action/impact-story-using-innovative-statistical-approaches-provide-most-reliable-treatment-outcomes#:~:text=Science%20in%20Action-,Impact%20Story%3A%20Using%20innovative%20statistical%20approaches%20to%20provide%20the%20most,information%20to%20patients%20and%20clinicians&text=Using%20Bayesian%20hierarchical%20models%2C%20CDER,affect%20different%20groups%20of%20patients

[ICH] International Council for Harmonisation. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials: ICH harmonised guideline. E9(R1). Adopted on November 20, 2019. Accessed November 07, 2022. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf

Inker LA, Schmid CH, Tighiouart H, et al; CKD-EPI Investigators. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29. https://doi.org/10.1056/NEJMoa1114248.

[JASSO] Japan Society for the Study of Obesity. 日本肥満学会. 肥満症診療ガイドライン 2022. 東京ライフサイエンス出版 [Guideline for the management of obesity disease 2022]. *Tokyo: Life Science Publishing*. 2022:167-168.

Kolotkin RL, Ervin CM, Meincke HH, et al. Development of a clinical trials version of the Impact of Weight on Quality of Life-Lite questionnaire (IWQOL-Lite clinical trials version): results from two qualitative studies. *Clin Obes*. 2017;7(5):290-299. https://doi.org/10.1111/cob.12197

Kolotkin RL, Williams VSL, Ervin CM, et al. Validation of a new measure of quality of life in obesity trials: Impact of Weight on Quality of Life-Lite clinical trials version. *Clin Obes*. 2019;9(3):e12310. https://doi.org/10.1111/cob.12310

Ma C, Shen X, Qu Y, Du Y. Analysis of an incomplete binary outcome dichotomized from an underlying continuous variable in clinical trials. *Pharm Stat*. 2022;21(5):907-918. https://doi.org/10.1002/pst.2204

Maruish ME, editor. *User's Manual for the SF36v2 Health Survey*. 3rd ed. Quality Metric Incorporated; 2011

PHUSE. Analysis and displays associated with adverse events: focus on adverse events in Phase 2-4 clinical trials and integrated summary documents. Published February 03, 2017. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+and+Displays+Associated+with+Adverse+Events+Focus+on+Adverse+Events+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary.pdf

PHUSE. Analysis and displays associated with demographics, disposition, and medications. Published March 02, 2018. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Demographics,+Disposition+and+Medication+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary+Documents.pdf

PHUSE. Analyses and displays associated with laboratory analyte measurements in Phase 2-4 clinical trials and integrated submission documents – update to recommendations. Published

August 19, 2022. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Safety+Analytics/WP068.pdf

PHUSE. Analyses and displays associated with measures of central tendency – focus on vital sign, electrocardiogram, and laboratory analyte measurements in Phase 2-4 clinical trials and integrated submission documents. Published October 10, 2013. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Measures+of+Central+Tendency-+Focus+on+Vital+Sign,+Electrocardiogram+%26+Laboratory+Analyte+Measurements+in+Phase+2-4+Clinical+Trials+and+Integrated+Submissions.pdf

PHUSE. Analyses and displays associated with outliers or shifts from normal to abnormal: focus on vital signs, electrocardiogram, and laboratory analyte measurements in Phase 2-4 clinical trials and integrated summary documents. Published September 10, 2015. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Outliers+or+Shifts+from+Normal+To+Abnormal+Focus+on+Vital+Signes+%26+Electrocardiogram+%26+Laboratory+Analyte+Measurements+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary.pdf

Röver C, Bender R, Dias S, et al. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Res Synth Methods*. 2021;12(4):448-474. https://doi.org/10.1002/jrsm.1475

Wang B, Du Y. Improving the mixed model for repeated measures to robustly increase precision in randomized trials. Published October 18, 2021. Accessed January 24, 2024. https://arxiv.org/abs/2110.09645

Ye T, Shao J, Yi Y, Zhao Q. Toward better practice of covariate adjustment in analyzing randomized clinical trials. *J Am Stat Assoc*. 2022;118(544):2370-2382. https://doi.org/10.1080/01621459.2022.2049278

Ye T, Bannick M, Yi Y, Shao J. Robust variance estimation for covariate-adjusted unconditional treatment effect in randomized clinical trials with binary outcomes. *Stat Theory Relat Fields.* 2023;7(2):159-163. https://doi.org/10.1080/24754269.2023.2205802

Signature Page for VV-CLIN-117489 v3.0

Approval	PPD 26-Jun-2025 03:26:13 GMT+0000

Signature Page for VV-CLIN-117489 v3.0