J1I-MC-GZBF Statistical Analysis Plan V4.0

A Phase 2 Study of Once-Weekly LY3437943 Compared with Placebo in Participants Who Have Obesity or Are Overweight with Weight-Related Comorbidities

NCT04881760

Approval Date: 15-Dec-2022

Statistical Analysis Plan (J1I-MC-GZBF): A Phase 2 Study of Once-Weekly LY3437943 Compared with Placebo in Participants Who Have Obesity or Are Overweight with Weight-Related Comorbidities

Protocol Title: A Phase 2 Study of Once-Weekly LY3437943 Compared with Placebo in Participants Who Have Obesity or Are Overweight with Weight-Related Comorbidities

Protocol Number: J1I-MC-GZBF

Compound Number: LY3437943

Short Title: Effect of LY3437943 versus Placebo in Participants Who Have Obesity or Are Overweight

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana USA 46285

Regulatory Agency Identifier Number(s)

Registry	ID
IND	154659

Document ID: VV-CLIN-076021

Table of Contents

List of Abbreviations and Terms

Term	Definition
AC	assessment committee
AE	adverse event
AESI	adverse event of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANHECOVA	analysis of heterogeneous covariance
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
bpm	beats per minute
CI	confidence interval
CEC	clinical endpoint committee
CN	conventional
COVID-19	Coronavirus disease 2019
CRF	case report form
CSR	clinical study report
CV	cardiovascular
DILI	Drug-Induced Liver Injury
EAS	Efficacy Analysis Set
ECG	electrocardiogram
eCRF	electronic case report form
eGFR	estimated glomerular filtration rate

Term	Definition
FAS	Full Analysis Set
GI	gastrointestinal
HLT	High Level Term
ICE	intercurrent event
ICH	International Council for Harmonisation
HbA1c	hemoglobin A1c
LLT	Lowest Level Term
LS	least squares
LY	LY3437943
MACE	major adverse cardiovascular events
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed-effect model repeated measures
PD	pharmacodynamic
PK	pharmacokinetic
PT	Preferred Term
QTcF	Fredericia's corrected QT interval
QW	once weekly
SAE	serious adverse event
SAP	statistical analysis plan
SC	subcutaneous
SD	standard deviation
SI	System International
SMQ	Standardised Medical Query

Term	Definition
SOC	System Organ Class
SS	Safety Analysis Set
T2D	type 2 diabetes
TBL	total bilirubin
TEAE	treatment-emergent adverse event
UACR	urine albumin-to-creatinine ratio
ULN	upper limit of normal

Version history

This SAP for Study J1I-MC-GZBF (GZBF) is based on the protocol dated 21 May 2021.

SAP Version	Approval Date	Change	Rationale
1.0	03-Dec-2021	Not Applicable	Original version
2.0	07-Apr-2022	• Remove the word "dulaglutide" from the SAP. • Added "Excludes participants discontinuing study drug due to inadvertent enrollment" to the definition of analysis set. • Updated ANCOVA to ANHECOVA throughout the SAP. • Updated definition of baseline to make it clearer and more concise. • Added CCI [redacted] as AESI in Section 4.6.	• "Dulaglutide" was not the active comparator of the study and got included by mistake • This is aligned with the new guidance on inadvertent enrollment. • Use ANHECOVA to achieve the efficiency gain. • Baseline definition is slightly different between efficacy analysis set and other analysis sets. • Per GPS recommendation.
3.0	12-Jul-2022	• Make clear the analysis set for analysis regarding safety parameters. • Add an alternative hybrid estimand as a sensitivity analysis.	• Make clear to use safety analysis set for analysis regarding safety parameters. • To support interactions with regulatory agency on conversations about hybrid estimand.

SAP Version	Approval Date	Change	Rationale
4.0	See date on Page 1	• The details and rationale of the third interim analysis was added to Section 4.8.	• The third interim analysis was requested by Lilly senior management to help support end of phase 2 interactions with regulatory agencies and needs to be documented in the SAP.
		• Add subgroup analyses based on quartiles of baseline BMI in Section 4.7.2.	• Explore the effect of QW LY3437943 in body weight change in population with different baseline BMI.
		• Added details about dose-pooling strategy in Kaplan-Meier estimation in Section 4.1.	• Similar with dose-pooling variant of MMRM, the dose-pooling strategy was considered in Kaplan-Meier so that treatment arms with the same pre-planned titration regimen has the same estimation during titration.
		• Added 'percentage of study participants who achieve ≥30% body weight reduction' and 'non-HDL cholesterol' as additional exploratory endpoints.	• Study team is also interested in exploring LY3437943's effects on the percentage of study participants who achieve ≥30% body weight reduction and changes in non-HDL cholesterol.
		• Analysis method for SF-36 in Section 4.7.1.2 is updated to be consistent with the one for eating	• Similar with the eating inventory, the analysis model for SF-36 should be ANHECOVA at the 24-week primary

SAP Version	Approval Date	Change	Rationale
		inventory.	analyses when there is only one post baseline measurement, and MMRM at the 48-week final analysis when there are two post baseline measurements.
		• The full model in the subgroup analyses in Section 4.7.2 is removed.	• The assumption for the full model is not appropriate.

1. Introduction

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Primary	
To demonstrate that LY3437943 1, 4, 8, or 12 mg is superior to placebo at 24 weeks from randomization for percent change in body weight.	Mean percent change in body weight
Secondary	
To demonstrate that LY3437943 1, 4, 8, or 12 mg is superior to placebo at 48 weeks from randomization for percent change in body weight.	Mean percent change in body weight

To demonstrate that LY3437943 1, 4, 8, or 12 mg is superior to placebo at 24 and 48 weeks from randomization for
• Body weight
• Waist circumference
 	
To assess safety and tolerability of study interventions
	• Percentage of study participants who achieve
	  ○ ≥5% body weight reduction
	  ○ ≥10% body weight reduction, and
	  ○ ≥15% body weight reduction
	• Mean change in body weight (kg)
	• Mean change in BMI (kg/m^2)
	• Mean change in waist circumference (cm)
	• Adverse events overall
	• Adverse events of special interest
	• Laboratory parameters
	• Electrocardiogram
	• Vital signs
	• Number of participants testing positive for anti-LY3437943 antibodies

Tertiary/Exploratory	

To investigate the effect QW LY3437943 1, 4, 8, or 12 mg versus placebo at various time points for:
• Body weight (at Weeks 24 and 48)
	• Percentage of study participants who achieve
	  ○ ≥20% body weight reduction
	  ○ ≥25% body weight reduction

Objectives	Endpoints
	○ ≥30% body weight reduction
• Blood pressure	• Mean change in ○ systolic BP (mmHg) measured by ABPM ○ diastolic BP (mmHg) measured by ABPM
• Heart rate	• Mean change in heart rate measured by ABPM
• Lipid parameters	• Mean change in fasting ○ total cholesterol ○ HDL cholesterol ○ LDL cholesterol ○ VLDL cholesterol, ○ non-HDL cholesterol and ○ triglycerides
• Glycemic control	• Mean change in ○ fasting glucose, and ○ HbA1c • Proportion of participants with incident T2D
• Mechanistic biomarkers	• Mean change in mechanistic biomarkers (see detailed list of parameters in Section 4.5)
• Patient-reported outcomes	• Proportion of participants with change in PGIS-Physical Function • Mean change in ○ SF-36v2 acute form domain scores, and ○ Eating Inventory domain scores
• To assess the PK of LY3437943 and potential participant factors that may influence its PK	• LY3437943 plasma concentrations
• To assess the relationship between LY3437943 dose and/or exposure and key efficacy and safety measures and potential participant factors that may influence these relationships	• Dose–response and concentration–response analyses for key efficacy and safety parameters

Abbreviations: ABPM = ambulatory blood pressure modeling; BMI = body mass index; BP = blood pressure; HbA1c = glycated hemoglobin A1c; HDL = high-density lipoprotein; LDL = low-density lipoprotein; PGIS = Patient Global Impression of Status; PK = pharmacokinetic(s); QW = weekly; SF-36v2 acute form = Short Form-36 Version 2 Health Survey acute form; T2D = type 2 diabetes mellitus; VLDL = very low-density lipoprotein.

Primary Estimand

The primary clinical question of interest is: What is the treatment difference in percent change in body weight from baseline after 24 weeks of treatment assuming all participants who meet the inclusion criteria would have completed the treatment period?

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

- Population: participants who meet the inclusion criteria. Further details can be found in Sections 5 and 9 of protocol J1I-MC-GZBF (a).
- Endpoint: percent change from baseline to 24 weeks in body weight.
- Treatment condition: the randomized treatment with allowance for down-titration based on GI tolerability.
- The ICE "permanent discontinuation of study drug" is handled by the hypothetical strategy and the potential outcome of interest is the response in the efficacy measurement if participants had adhered to the randomized treatment. There are no other defined ICEs. Down-titration will not be considered as ICEs for the definition of estimand in this study.
- Population-level summary: mean percent changes in body weight at Week 24.

Treatment effect of interest is thus defined as the difference in mean percent changes in body weight at Week 24 between QW LY3437943 and placebo.

Rationale for "efficacy estimand": This Phase 2 study aims to study the efficacy of LY3437943 under the ideal condition that all participants adhere to the randomized treatment.

Estimand(s) for Secondary Objectives

The same estimand for the primary objective will be used for the following efficacy endpoints for the secondary objectives:

- difference between LY3437943 and placebo in percent change from baseline in body weight at Week 48
- percentage of participants with 5% or higher, 10% or higher, and 15% or higher body weight loss from baseline at Weeks 24 and 48
- change from baseline in body weight (kg) at Weeks 24 and 48
- change from baseline in BMI (kg/m^2) at Weeks 24 and 48; and
- change from baseline in waist circumference (cm) at Weeks 24 and 48.

Unless specified otherwise, safety and tolerability assessments will be guided by an estimand comparing safety of LY3437943 doses with placebo for the whole study period (the treatment period plus safety follow-up period) irrespective of adherence to study intervention for all study population (including inadvertent enrollment).

Supplemental Estimand(s) for Primary Efficacy Endpoint

An alternative estimand, called "hybrid estimand", is the mean treatment difference in percent change in body weight from baseline at Week 24 between LY3437943 and placebo in participants who meet the inclusion criteria with ICEs handled differently according to the reasons for the events:

- Category 1: The ICEs of permanent discontinuation of study drug due to reasons unlikely related to the efficacy/safety outcomes will be handled by the hypothetical strategy.
- Category 2: The ICEs of permanent discontinuation of study drug due to lack of efficacy before study treatment discontinuation will be handled by the hypothetical strategy
- Category 3: All other ICEs will be handled by the treatment policy strategy. With treatment policy strategy, the occurrences of "permanent discontinuation of study drug" is considered irrelevant in defining the treatment effect of interest. The value for the response variable is used regardless of whether or not the ICE occurs.

Population-level summary: mean percent change in body weight at Week 24.

Rationale for "hybrid estimand": following ICH E9 (R1) guidance on estimand the study will collect informative treatment disposition reasons and ICEs will be handled separately according to the reasons of ICEs for this supplemental estimand. This supplemental estimand will also be applied to the efficacy endpoints for the secondary objectives listed in the Estimand(s) for Secondary Objectives portion of Section 1.1.

Further details can be found in Section 4.3.3.

Another alternative estimand, called "adherer estimand", is the treatment difference in percent change in body weight at Week 24 between LY3437943 and placebo in participants who would adhere to both LY3437943 and placebo.

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

- Population: a principal stratum including participants who meet the inclusion criteria and would adhere to both LY3437943 and placebo (note that this is a hypothetical population). We define that a participant adheres to the treatment if he/she has taken at least 75% of the study drug without permanent study drug discontinuation.
- Endpoint: percent change from baseline to Week 24 in body weight.
- Treatment condition: the randomized treatment with allowance for dose modification based on GI tolerability.
- Population-level summary: mean percent changes in body weight at Week 24.

Treatment effect of interest is thus defined as the difference in mean percent changes in body weight at Week 24 between QW LY3437943 and placebo.

Rationale for "adherer estimand": to study the efficacy of LY3437943 versus placebo among those participants could adhere to both treatments.

1.2. Study Design

Study GZBF is a 48-week, Phase 2, multicenter, randomized, double-blind, placebo-controlled study designed to examine the safety and efficacy of 4 dose levels of SC QW administered

LY3437943 compared with SC QW administered placebo in participants who have obesity (BMI 30 kg/m^2 or higher) or are overweight (BMI 27 kg/m^2 or higher and less than 30 kg/m^2) with weight-related comorbidities but without T2D. The primary objective will be the effect of LY3437943 on percent change in body weight at 24 weeks. Four maintenance doses of LY3437943 will be evaluated in the trial: 1, 4, 8, and 12 mg. Dose escalation to improve GI tolerability will occur in certain treatment groups up to Week 12 by increasing the volume of administered study drug (or placebo). For maintenance doses equal to or greater than 4 mg, the initial dose will be 2 or 4 mg followed by additional escalation steps. For 2 of the maintenance dose arms (4 and 8 mg), participants will be randomly assigned into 2 subgroups, with different dose escalation schemes. This is described in detail in Section 4.1.1 of the protocol. Study participants will be randomly assigned in a 2:1:1:1:1:2:2 ratio (LY3437943 1 mg QW, LY3437943 4 mg [2 mg] QW, LY3437943 4 mg QW [4 mg], LY3437943 8 mg QW [2 mg], LY3437943 8 mg QW [4 mg], LY3437943 12 mg [2 mg] QW, placebo QW) and stratified by sex and BMI 36 kg/m^2 or higher. Study GZBF will consist of 3 periods: a 6-week screening period, a 48-week double-blind, placebo-controlled treatment period (consisting of an up to 12-week dose-escalation period and a 36-week maintenance period), and a 4-week safety follow-up period.

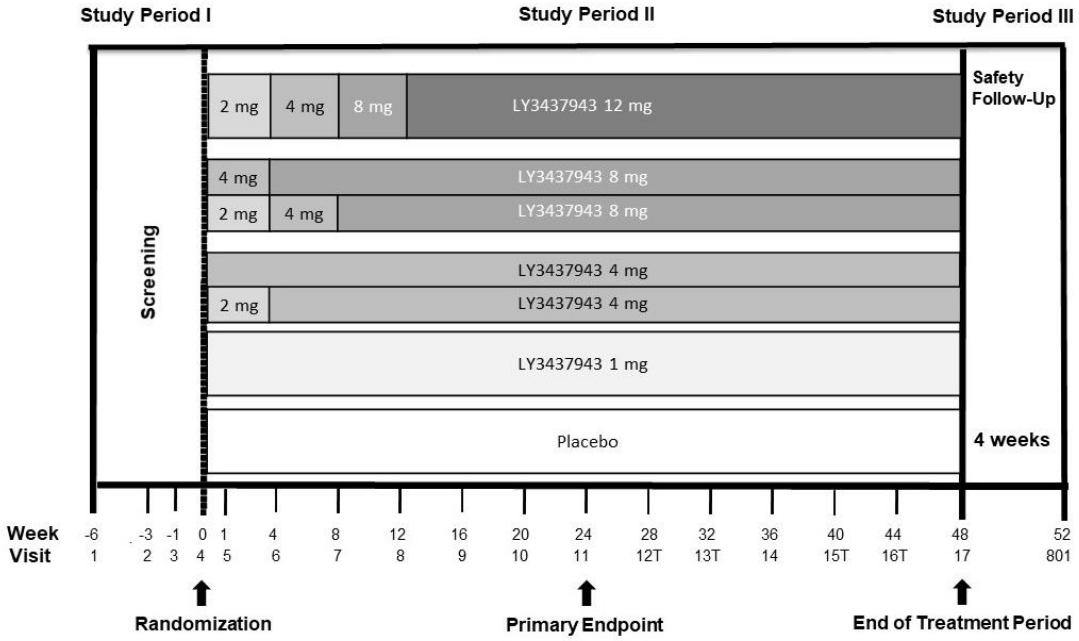

Figure GZBF.1.1. Illustration of study design for clinical protocol J1I-MC-GZBF.

2. Statistical Hypothesis

The primary objective is to demonstrate that LY3437943 1, 4, 8, or 12 mg administered SC QW is superior to placebo with respect to percent change in body weight from baseline to Week 24, in participants without T2D who have obesity or are overweight with weight-related comorbidities. Thus, the null hypothesis to be tested in relation to the primary estimand is as follows:

- null hypothesis: each dose level of LY3437943: 1, 4, 8, or 12 mg is not different from placebo with respect to percent change in body weight from baseline to Week 24.

2.1. Multiplicity Adjustment

No adjustment for multiplicity will be performed.

3. Analysis Sets

For the purposes of analysis, the following analysis sets are defined:

Participant Analysis Set	Description
Entered	All participants who sign the informed consent form.
Randomized	All participants who are randomly assigned to a treatment arm.
Efficacy Analysis Set (EAS)	Data obtained during treatment period from all randomly assigned participants. Excludes participants discontinuing study drug due to inadvertent enrollment and data after discontinuation of study drug. Participants will be analyzed according to the treatment group to which they were randomly assigned.
Full Analysis Set (FAS)	Data obtained during treatment period from all randomly assigned participants who take at least 1 dose of double-blind study treatment, excluding participants discontinuing study drug due to inadvertent enrollment, regardless of adherence to study drug. Participants will be analyzed according to the treatment group to which they were randomly assigned.
Safety Analysis Set (SS)	Data obtained during treatment period plus safety follow-up period from all randomly assigned participants who take at least 1 dose of double-blind study treatment, regardless of adherence to study drug. Participants will be analyzed according to the treatment group to which they were randomly assigned.

4. Statistical Analyses

4.1. General Considerations

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly) or its designee. Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (eg, few events to justify conducting an analysis). Additional analyses of the data may be conducted as deemed appropriate.

Unless otherwise noted, tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and the CI will be calculated at 95% 2-sided. All tests of interactions between treatment groups and other factors will be conducted at a 2-sided alpha level of 0.10.

Unless otherwise specified, the efficacy analysis will be conducted using EAS and the safety analysis will be conducted using SS.

Unless stated otherwise, statistical summaries and analyses will be conducted based on planned randomized treatment group: placebo, LY 1.0 mg, LY 4.0 mg (2.0 mg), LY 4.0 mg (4.0 mg), LY 8.0 mg (2.0 mg), LY 8.0 mg (4.0 mg), and LY 12.0 mg (2.0 mg), regardless of the actual treatment(s) received by the participant due to any dose modification. The evaluation of the efficacy and safety endpoints for LY 4.0 mg and LY 8.0 mg compared with placebo or dulaglutide will be made by pooling 2 dose escalation regimens, ie, combine LY 4.0 mg (2.0 mg) and LY 4.0 mg (4.0 mg) as LY 4.0 mg (pooled), and combine LY 8.0 mg (2.0 mg) and LY 8.0 mg (4.0 mg) as LY 8.0 mg (pooled). Therefore, the statistical comparisons will be made between each of LY3437943 treatment groups: LY 1.0 mg, LY 4.0 mg (2.0 mg), LY 4.0 mg (4.0 mg), LY 4.0 mg (pooled), LY 8.0 mg (2.0 mg), LY 8.0 mg (4.0 mg), LY 8.0 mg (pooled), LY 12.0 mg (2.0 mg) and placebo.

The primary estimand (a precise definition of the treatment effect to be estimated) of interest in comparing efficacy of LY3437943 doses with placebo is the “efficacy estimand” (Section 1.1). The primary efficacy assessment, guided by the “efficacy estimand” will be conducted using the EAS (Section 3). A restricted maximum likelihood-based, dose-pooling, MMRM analysis will be used to analyze continuous longitudinal variables. All the longitudinal observations at each scheduled postbaseline visit will be included in the analysis. This dose-pooling variant of MMRM replaced the fixed effect of treatment group with that of the preplanned time-varying treatment doses at each time point in an effort to derive a more efficient estimator when the same treatment regimen over time is shared across the treatment groups due to preplanned titration (Qu et al. 2021). For example, in LY 4.0 mg (2.0 mg) and LY 8.0 mg (2.0 mg) arms, the doses administered for the first 4 weeks are both 2.0 mg. Therefore, pooling all the patients in these 2 groups provides a more efficient estimation in the effect of 2.0 mg which is also anticipated to increase the estimation precision for the later administered doses. The model for the analysis of the primary efficacy endpoint of percent change from baseline in body weight will include the fixed effects of treatment doses planned in the titration scheme: placebo, LY 1.0 mg, LY 2.0 mg, LY 4.0 mg (2.0 mg), LY 4.0 mg (4.0 mg), LY 8.0 mg (2.0 mg), LY 8.0 mg (4.0 mg), and LY 12.0 mg (2.0 mg), stratification factors including gender (males, females) and baseline BMI stratum (less than 36, greater than or equal to 36), and continuous, fixed covariate of baseline value, all nested within visits. An unstructured covariance structure will be used to model the within-participant errors. Significance tests will be based on LS means and robust “Huber-

White" standard errors. If this analysis fails to converge, the following covariance structures will be tested in order:

- autoregressive, and compound symmetry. The first covariance structure that converges will be used.

Unless specified otherwise, safety assessments will be guided by an estimand comparing safety of LY3437943 doses with placebo irrespective of adherence to study drug. Thus, safety analyses will be conducted using the SS.

For continuous measures, summary statistics will include sample size, mean, SD, median, minimum, and maximum for both the actual and the change from baseline measurements. LS means and standard errors derived from the analysis models will also be displayed for the change from baseline measurements. Treatment comparisons will be displayed showing the treatment difference LS means and the 95% CIs for the treatment differences, along with the p-values for the treatment comparisons. All baseline measures will be analyzed using an analysis of variance model that has treatment group as the model terms.

For categorical measures, summary statistics will include sample size, frequency, and percentages. A logistic regression model will be used to examine the treatment difference in binary efficacy outcomes with missing endpoints imputed. Fisher's exact test or Pearson's chi-square test will be used for treatment comparisons in other categorical outcomes.

For laboratory values, both CN and SI units will be presented. For body weight, kilogram (kg) will be presented.

Summary statistics for discrete count measures will include sample size, mean, SD, median, minimum, and maximum.

Kaplan-Meier method will be used for estimation of cumulative event-free survival rates over time, and Cox proportional hazards regression analysis will be used to compare hazards rates among treatments. In Kaplan-Meier estimation, the dose-pooling strategy will be used: subjects in the pre-specified analysis set with the same pre-planned titration history will be pooled together to calculate the number of events and the total number of subjects at risk.

Unless otherwise specified, baseline is defined as the last nonmissing measurement prior to first dosing of the study drug, or prior to the randomization date for the subjects who are randomized but never dosed.

End of study participation for a participant will be the earliest of date of death, date of withdrawal from further participation in the study, or date of safety follow-up visit (Visit 801). For participants considered to be lost-to-follow-up, end of study participation will be the date of lost-to-follow-up reported by the investigator. Participant data included in the database after the last date of study participation (including safety follow-up period) will be excluded from statistical analysis.

Statistical treatment comparisons will only be performed between LY3437943 and placebo. Since the trial is not adequately powered to detect difference among LY3437943 doses, comparisons among LY3437943 doses will not be performed unless otherwise specified.

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.

4.2. Participant Dispositions

A listing and summary of study disposition for all randomized participants will be provided at the primary database lock and final database lock, respectively. Frequency counts and percentages of all participants screened, randomized, and receiving at least 1 dose of study drug will be presented by treatment groups and dose escalation subgroups. A listing and summary of randomized participants not receiving study drug will be provided. All participants who discontinue the study and/or study drug will be identified, and the extent of their participation in the study will be reported. If known, a reason for their discontinuation will be given. The primary reasons for discontinuation will be listed and summarized by treatment groups and dose escalation subgroups. Kaplan-Meier plots of time to premature study discontinuation, premature study treatment discontinuation, and premature study treatment discontinuation due to AE will be provided based on all randomized population. Time-to-event analyses of premature study treatment discontinuation, study drug discontinuation due to AE, and study discontinuation may be conducted.

Details about the demographic and baseline characteristics, historical illnesses, and preexisting conditions, concomitant medications, treatment compliance, and important protocol deviations can be found in Appendix 1 (Section 7.1) through Appendix 5 (Section 7.5), respectively.

4.3. Primary Endpoint Analysis

The primary efficacy assessment, guided by the "efficacy estimand", will be conducted using the EAS. For the "efficacy estimand", the hypothetical strategy is used to handle the ICE (permanent discontinuation of study drug), so only data collected before the occurrence of any such ICEs will be used in the MMRM estimation (Section 4.1). Through the MMRM, the potential efficacy measures (after the ICEs) had participants not experienced ICEs will be implicitly imputed. To confirm efficacy of LY3437943 with adequate statistical power, the evaluation of the efficacy and safety endpoints for LY 4.0 mg and LY 8.0 mg compared with placebo will be made by pooling 2 dose escalation regimens, ie, combine LY 4.0 mg (2.0 mg) and LY 4.0 mg (4.0 mg) as LY 4.0 mg (pooled), and combine LY 8.0 mg (2.0 mg) and LY 8.0 mg (4.0 mg) as LY 8.0 mg (pooled). The primary efficacy comparison will be based on the contrast between each treatment group of LY3437943 and placebo at Week 24 (Visit 13) from the MMRM analysis of percent change from baseline in body weight using the EAS (Section 4.1). The analysis model and selection of covariance structure is described in Section 4.1. Treatment comparisons will be performed for the primary objective at the full significance level of 0.05.

4.3.1. Definition of Endpoint

The primary efficacy measure will be percent change in body weight from baseline (Week 0) to Week 24. The percent change in body weight at each nominal visit is defined as:

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

4.3.2. Main Analytical Approach

Percent change from baseline in body weight will be analyzed using the MMRM model for the "efficacy estimand" as described in Section 4.1.

4.3.3. Analyses for Supplemental Estimand(s)

The analysis related to a supplemental estimand, "hybrid estimand" (Section 1.1), will be conducted using data in the randomized population.

Hybrid estimand is defined as the treatment difference in the mean percent change in body weight from baseline at Week 24 between LY3437943 and placebo for the study target population with ICEs handled differently according to the reasons of the events as follows:

- Category 1: The ICEs of permanent discontinuation of study drug due to reasons unlikely related to the efficacy/safety outcomes will be handled by the hypothetical strategy.
- Category 2: The ICEs of permanent discontinuation of study drug due to lack of efficacy before study treatment discontinuation will be handled by the hypothetical strategy.
- Category 3: All other ICEs will be handled by the treatment policy strategy.

In this study, following ICH E9 (R1) guidance, a plan was made to collect informative treatment disposition reasons, through eCRF, for why data intended for collection are missing and classify them into categories 1 through 3 as shown in Table GZBF.4.1 (ICH 2019).

Table GZBF.4.1. Treatment Disposition Reasons

Disposition Reason	Associated Sub-Categories	Category
Adverse event		3
Protocol deviation	Due to epidemic/pandemic	1
	Other	3
Pregnancy		3
Lack of efficacy		2
Other		3
Withdrawal by subject	Concern about study procedures/perceived risks	3
	Scheduling conflicts	1
	Subject is moving or has moved	1
	Personal issue unrelated to trial	1
	Due to epidemic/pandemic	1
	Other (option to include a specify field)	3
Physician decision	Concern about study procedures/perceived risks	3
	Scheduling conflicts	1
	Subject is moving or has moved	1
	Due to epidemic/pandemic	1
	Other (option to include a specify field)	3
Study terminated by sponsor		1
Site terminated by sponsor		1
Study terminated by IRB/ERB		1

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

To estimate the "hybrid estimand," multiple imputation will be used to impute the corresponding missing potential outcome according to the missingness patterns (Table GZBF.4.2) with ICEs handled differently according to the reasons of the events.

When participants have missing values without ICEs, the missing values will be imputed using data from participants who do not have ICE or missing values.

Percent change from baseline in body weight will then be analyzed using ANHECOVA with terms of treatment group: placebo, LY 1.0 mg, LY 4.0 mg (2.0 mg), LY 4.0 mg (4.0 mg), LY 8.0 mg (2.0 mg), LY 8.0 mg (4.0 mg) and LY 12.0 mg (2.0 mg), stratification factors including gender (female, male) and baseline BMI stratum (less than 36, greater than or equal to 36), and continuous covariate of baseline value, along with the interaction between treatment group and the stratification factors and that between treatment group and the baseline value. The heteroscedasticity is also considered by introducing treatment specific variance estimates. This formulation has been shown to provide the optimal efficiency gain regardless of randomization schemes within the class of linear model adjustment (Ye et al. 2009).

The evaluation for LY 4.0 mg and LY 8.0 mg compared with placebo will be made by pooling 2 dose escalation regimens, ie, combine LY 4.0 mg (2.0 mg) and LY 4.0 mg (4.0 mg) as

LY 4.0 mg (pooled), and combine LY 8.0 mg (2.0 mg) and LY 8.0 mg (4.0 mg) as LY 8.0 mg (pooled).

Table GZBF.4.2. Strategy to Handle ICE and Missingness for Hybrid Estimand

ICE	Strategy to Handle ICE	Assumptions for Missingness	Methods to Handle Missing Values at Endpoint
Category 1: Treatment discontinuation due to reasons unlikely related to efficacy/safety outcome	Hypothetical	MAR	Data collected after the ICE will be set to missing. Missing values will be imputed using all nonmissing data (excluding data collected after ICEs) from the same treatment arm.
Category 2: Treatment discontinuation due to lack of efficacy	Hypothetical	MAR	Data collected after the ICE will be set to missing. Missing values will be imputed using all nonmissing data (excluding data collected after ICEs) from the same treatment arm.
Category 3: All other treatment discontinuations	Treatment policy	MNAR Considers that these participants could not adhere to their assigned treatment and may not benefit from the assigned treatment.	Missing values will be imputed using participants in the same treatment arm with similar ICEs but nonmissing values (retrieved dropout imputation). In cases where there are not enough retrieved dropouts to provide a reliable imputation model, will impute the missing data using the jump-to-reference (placebo) imputation approach. For Study GZBF, the likelihood of sufficient retrieved dropouts is assumed low.

Abbreviations: ICE = intercurrent event; MAR = missing at random; MNAR = missing not at random.

A sensitivity analysis may be conducted related to an alternative version of a hybrid estimand with the details provided below. The purpose of such a sensitivity analysis is to explore the best definition and handling strategy for the hybrid estimand in support of the compound development for chronic weight management. The difference from the aforementioned hybrid estimand mainly lies in the strategy to handle ICEs:

- Category 1: The ICEs of permanent discontinuation of study treatment due to reasons unlikely related to the efficacy/safety outcomes (i.e., personal reasons, scheduling conflicts, relocation, pandemic, geographic conflicts, etc.) will be handled by the hypothetical strategy. Generally, these ICEs will also cause participants to discontinue from the study. The potential outcome of the response if participants would continue the study treatment will be of interest. This is of best clinical interest as in real life — either these ICEs do not prevent participants from taking the medication or these ICEs (pandemic or geographic conflicts) do not represent the situation at normal time.
- Category 2: The ICEs of permanent discontinuation of study treatment due to adverse events or lack of efficacy will be handled by the treatment policy strategy. The potential outcome is the response after stopping taking the study treatment from the time of treatment discontinuation without the use of other anti-obesity medications.

- Category 3: The ICEs of the use of anti-obesity medication post permanent treatment discontinuation will be handled by hypothetical strategy. The potential outcome of the response without using the anti-obesity medication will be the interest.
- Category 4: All other ICEs, i.e., the ICEs of permanent discontinuation of study treatment due to all other reasons, will be handled by the treatment policy strategy. The potential outcome is the response after stopping taking the study treatment from the time of treatment discontinuation without the use of other anti-obesity medications.

The strategies for imputing missing data for different categories of ICEs are as follows, in line with the corresponding strategies.

- Category 1: data collected after these ICEs are excluded. The resulting missing values can be imputed using those participants who adhere to the study medication in the same treatment group under the assumption of Missing At Random (MAR).
- Categories 2 and 4: data collected after these intercurrent events will be used. Additional missing values will be imputed by using the observations for those who discontinue study treatment due to these corresponding ICEs and who have data collected post these ICEs (retrieved dropout approach) or by jump to reference (placebo imputation) under the assumption of Missing Not At Random (MNAR).
- Category 3: data after initiation of the ICEs will be excluded. Since anti-obesity medication other than LY3437943 is prohibited until the participants discontinue the study treatment, we assume that the ICEs of such treatment discontinuation associated with the use of anti-obesity medication belong to the Categories 2 or 4, e.g., study treatment discontinuation due to lack of efficacy. The resulting missing values will be imputed using the strategies as defined in Category 2 and 4.

The alternative methods of handling ICEs and imputing missing values are summarized in Table GZBF.4.3.

Table GZBF.4.3. Strategy to Handle ICE and Missingness for Hybrid Estimand

Type of ICE	Strategy to handle ICE	Potential outcome	Handling data after ICE	Missing data imputation
Category 1	Hypothetical	The outcome if participants would adhere to the study medication	To discard	Imputing using the observed data up to ICEs from the same treatment group, based on the MAR assumption
Category 2	Treatment policy	The outcome as the result of ICE	To use	Imputing using retrieved dropouts or placebo arm,
Category 3	Hypothetical	The outcome if participants would not	To discard	

Type of ICE	Strategy to handle ICE	Potential outcome	Handling data after ICE	Missing data imputation
		take anti-obesity medication		based on MNAR assumption
Category 4	Treatment policy	The outcome as the result of ICE	To use	

Abbreviations: ICE = intercurrent event; MAR = missing at random; MNAR = missing not at random.

The analysis related to "adherer estimand" will be conducted using data in the EAS. "Adherer estimand" is the treatment difference in percent change in body weight from baseline at Week 24 between LY3437943 and placebo in participants who would adhere to both LY3437943 and placebo. We define that a participant adheres to the treatment if he/she has taken at least 75% of the study drug without permanent study drug discontinuation. A multiple imputation approach proposed by Luo et al. (2021) will be used. The data used for the imputation may include the baseline covariates, intermediate outcomes/variables, adherence status and the final outcome of percent change in body weight from baseline at Week 24. Such data for each treatment group is used to impute the potential outcomes (including intermediate and final outcomes, and adherence status) for participants assigned to other treatment groups.

After imputation, for each participant the potential outcomes for the final outcome of percent change in body weight from baseline at Week 24 and adherence status under all the treatments are available. Then, the mean response for each treatment group can be estimated by simply taking the average of the potential final outcomes for the adherers. The estimation of the treatment difference for the "adherer estimand" can be subsequently derived by taking the difference between the calculated mean responses. The inference is provided based on variance estimated by bootstrap.

4.4. Secondary Endpoints Analysis

Unless otherwise specified, secondary efficacy analyses will be conducted for EAS. Decision will be guided by the 2-sided p-values in each objective (see Table GZBF.4.4).

Table GZBF.4.4. Secondary Measures Not Controlled for Type 1 Error

Objectives	Relative to the efficacy measure	Analysis conducted in a manner similar to	Additional information
LY3437943 1, 4, 8, or 12 mg is superior to placebo at Week 48 from randomization (Week 0)	Mean percent change in body weight	MMRM model in Section 4.3.2	Same independent variables as in the MMRM model in Section 4.3.2. LSM estimates will be plotted by treatment through 48 weeks.
LY3437943 1, 4, 8, or 12 mg is superior to placebo at various visits from randomization (Week 0)	Percentage of study participants who achieve ≥5% body weight reduction ≥10% body weight reduction ≥15% body weight reduction	Logistic regression model with multiple imputation assuming ignorable missingness (Ma et al. 2022)	Use treatment group, stratification factors and continuous baseline value as covariates. Missing postbaseline continuous-valued body weight data are imputed first within each treatment arm before deriving the binary outcome. Plot of proportion of patients achieving ≥5%, ≥10%, and ≥15% body weight reduction will also be provided. Kaplan-Meier plots of time to initially achieve ≥5%, ≥10%, and ≥15% body weight reduction will be provided.
	Mean change in body weight (kg)	MMRM model in Section 4.3.2 with change in body weight (kg) as the response variable	Same independent variables as in the MMRM model in Section 4.3.2. LSM estimates will be plotted by treatment through 48 weeks.
	Mean change in BMI (kg/m^2)	MMRM model in Section 4.3.2 with change in BMI (kg/m^2) as the response variable	Use baseline BMI (kg/m^2) as a covariate. LSM estimates through 48 weeks will be plotted by study treatment.
	Mean change in waist circumference (cm)	MMRM model in Section 4.3.2 with change in waist circumference (cm) as the response variable	Use baseline waist circumference as a covariate. LSM estimates through 48 weeks will be plotted by study treatment.

Abbreviations: BMI = body mass index; LSM = least squares mean; MMRM = mixed model for repeated measures.

4.4.1. Analyses for Supplemental Estimand(s)

Mean percent change in body weight at Week 48 will also be analyzed by ANHECOVA guided by the hybrid estimand (similar analysis in Section 4.3.3). Missing body weight values at Week 48 will be imputed by the multiple imputation method described in Section 4.3.3.

Analysis guided by "adherer estimand" might also be conducted for these endpoints (similar analysis in Section 4.3.3).

4.5. Tertiary/Exploratory Endpoints Analysis

Unless otherwise specified, exploratory analyses will be conducted for EAS. Decision will be guided by the 2-sided p-values in each objective.

Table GZBF.4.5. Exploratory Measures

Objectives	Relative to the efficacy measure	Analysis conducted
Compare LY3437943 1, 4, 8, or 12 mg with placebo at various visits from randomization (Week 0)	Percentage of study participants who achieve • ≥20% body weight reduction • ≥25% body weight reduction • ≥30% body weight reduction	Same logistic regression model in Table GZBF.4.2. Plot of proportion of patients achieving ≥20%, ≥25%, and ≥30% body weight reduction will also be provided. Kaplan-Meier plots of time to initially achieve ≥20%, ≥25%, and ≥30% body weight reduction will be provided.
Compare LY3437943 1, 4, 8, or 12 mg with placebo at Weeks 24 and 36 from baseline (Week -1)	Mean change in • systolic BP (mmHg) measured by ABPM • diastolic BP (mmHg) measured by ABPM	Use treatment policy estimand with all the data available in safety analysis set for MMRM model in Section 4.3.2. If using ANHECOVA model in Section 4.3.3, multiple imputation will be done by treatment group with all available data in safety analysis set.
	Mean change in heart rate measured by ABPM	Same above.
Compare LY3437943 1, 4, 8, or 12 mg with placebo at Weeks 24 and 48 from randomization (Week 0)	Mean change in the following fasting lipid parameters • total cholesterol • HDL cholesterol • LDL cholesterol • VLDL cholesterol • non-HDL cholesterol • triglycerides	MMRM model in Section 4.3.2 with baseline response as covariate and log transformation for the response variables. LSM estimates through 48 weeks will be plotted by study treatment.
Compare LY3437943 1, 4, 8, or 12 mg with placebo at Weeks 24, 36 and 48 from randomization (Week 0)	Mean change in • fasting glucose (mg/dL and mmol/L) • HbA1c (% and mmol/mol)	MMRM model in Section 4.3.2 with baseline response as covariate. LSM estimates through 48 weeks will be plotted by study treatment.
Compare LY3437943 1, 4, 8, or 12 mg with placebo at Weeks 24 and 48 from randomization (Week 0)	Proportion of participants with incident T2D	Logistic regression model in Table GZBF.4.1 with baseline HbA1c as covariate.
Compare LY3437943 1, 4, 8, or 12 mg with placebo at Weeks 24 and 48 from randomization (Week 0)	Mean change in mechanistic biomarkers: • fasting insulin • fasting C-peptide • total adiponectin • CCI [redacted] • leptin • intact proinsulin • HOMA2-IR • HOMA2-B • intact proinsulin/C-peptide ratio • intact proinsulin/insulin ratio • CCI [redacted]	• For endpoint biomarkers in the list below, use ANHECOVA model where multiple imputation will use only hypothetical strategy to deal with ICEs. • total adiponectin • CCI [redacted] • leptin • beta-hydroxybutyrate • uric acid • Apo B • Apo C3 • hsCRP

Objectives	Relative to the efficacy measure	Analysis conducted
	CCI • beta-hydroxybutyrate • free fatty acids • uric acid • Apo B • Apo C3 • hsCRP	For longitudinal biomarkers in the list below, MMRM model in Section 4.3.2 with baseline response as covariate. • fasting insulin • fasting C-peptide • HOMA2-IR • HOMA2-B • intact proinsulin • intact proinsulin/C-peptide ratio • intact proinsulin/insulin ratio • CCI • free fatty acids LSM estimates through 48 weeks will be plotted by study treatment. Log transformation of the response variables may be conducted.
Compare LY3437943 1, 4, 8, or 12 mg with placebo at Weeks 24 and 48 from randomization (Week 0)	Proportion of participants with change in PGIS-Physical Function	See Section 4.7.1.1.
Compare LY3437943 1, 4, 8, or 12 mg with placebo at Weeks 24 and 48 from randomization (Week 0)	Mean change in • SF-36v2 acute form domain scores • Eating Inventory domain scores	See Sections 4.7.1.2 and 4.7.1.3.

Abbreviations: ABPM = ambulatory blood pressure modeling; ANHECOVA = analysis of heterogeneous covariance; Apo B = apolipoprotein B; Apo C3 = apolipoprotein C-III; BP = blood pressure; CCI ; HbA1c = glycated hemoglobin A1c; HDL = high-density lipoprotein; HOMA2-B = homeostasis model assessment of beta-cell function; HOMA2-IR = homeostasis model assessment of insulin resistance; hsCRP = high sensitivity C-reactive protein; LDL = low-density lipoprotein; LSM = least squares mean; MMRM = mixed model for repeated measures; PGIS = Patient Global Impression of Status; SF-36v2 = Short Form 36 version 2 Health Survey; T2D = type 2 diabetes mellitus; VLDL = very low-density lipoprotein.

4.5.1. Pharmacokinetic and Pharmacokinetic/Pharmacodynamic Methods

Pharmacokinetic, PD, and PK/PD analysis are the responsibility of Lilly's PK/PD group.

LY3437943 concentration-time data will be summarized in the CSR. Exposure-response analysis between LY3437943 concentration and safety, pharmacology, and efficacy may be performed using population PK and population PK/PD nonlinear mixed-effects modeling techniques implemented on Nonlinear Mixed Effects Modeling software. Additionally, the impact of intrinsic and extrinsic factors (such as age, weight, sex, renal, and hepatic functions) on PK and/or PD parameters may be evaluated.

4.5.2. Bayesian Analyses for Dose-Response

The longitudinal dose-response model as proposed by Qu et al. (2019) will be adapted here. This model also considers the preplanned dose changes that occur during the titration period. Let $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ be the doses a participant has planned to take and $t_c = (t_{c1}, t_{c2}, \dots, t_{cm})$ be the corresponding times when the dose changes where t_{ci} indicates the time for dose to change from θ_i to θ_{i+1}. Therefore, the mean function of the parameter of interest at time t is modelled by:

$$f_{\theta,t_c}(t) = f(t;\theta_1) + \sum_{i=1}^{m-1}[f(t-t_{ci};\theta_{i+1}) - f(t-t_{ci};\theta_i)]I(t > t_{ci}),$$

where I(X) is the indicator function that takes value 1 when the condition X holds. The $f(t;\theta)$ is defined such that

$$f(t;\theta) = \frac{\lambda(\theta)(1-e^{-k(\theta)t})}{1-e^{-k(\theta)d}},$$

where d is the maximum duration of the treatment period in weeks ($d = 24$ for primary analyses while $d = 36$ for final analysis), $\lambda(\theta)$ is the dose-response function for the maximum response at dose θ and $k(\theta)$ is dose $\theta's$ rate parameter. This formulation of the mean function $f(t;\theta)$ was introduced by Fu and Manner (2010) to characterize the change from baseline over time in a continuous outcome that could be approximated with a pattern of exponential decay. It assumes a monotone time profile with the maximum effect reached at time d. The longitudinal data $Y_{\theta,t_c,jt}$ for participant j at time t with titration scheme (θ, t_c) will be fitted by adding the error terms to the mean function $f_{\theta,t_c}(t)$ where

$$Y_{\theta,t_c,jt} = f_{\theta,t_c}(t) + \frac{s_j(1-e^{-k(\theta_1)t})}{1-e^{-k(\theta_1)d}} + \epsilon_{jt},$$

$s_j \sim N(0,\sigma_s^2)$ and $\epsilon_{jt} \sim N(0,\sigma^2)$ are independent, denoting between-subject variation and within-subject variation respectively. Given s_j, $Y_{\theta,t_c,jt} \sim N(f_{\theta,t_c}(t) + \frac{s_j(1-e^{-k(\theta_1)t})}{(1-e^{-k(\theta_1)d})}, \sigma^2)$.

For characterizing the body weight change, we use a different formulation of $f_{\theta,t_c}(t)$ that showed better fitting in the analyses of historical trials and simulation studies while keeping all other elements unchanged. In the new formulation,

$$f_{\theta,t_c}(t) = f(t;\theta_1) + \sum_{i=1}^{m-1} h(t-t_{ci})[f(t;\theta_{i+1}) - f(t;\theta_i)]I(t > t_{ci}),$$

Where $h(t-t_{ci}) = \frac{(1-e^{-\tau(t-t_{ci})})}{1-e^{-\tau d}}$. The parameter τ controls the rate of change in the body weight when dose changes. The dose-maximum response function $\lambda(\theta)$ is provided below:

The dose-maximum response function $\lambda(\theta)$ is provided below:

- body weight.

A power model is assumed where

$$\lambda(\theta) = a + b * \theta^{\gamma}$$

γ is a sigmoidicity parameter indicating shape or steepness of dose response.

Other dose-response models for $\lambda(\theta)$ may be explored if the aforementioned dose-response models do not fit the data well, eg, Simple Normal Dynamic Linear Modeling. Additionally, the different formulation of $f_{\theta,t_c}(t)$ may also be considered if the current specification does not provide an adequate fit to the data.

The estimation of those parameters will be carried out in a Bayesian framework assuming noninformative priors for the hyperparameters in the model as follows:

$$\begin{cases} k(\theta) \sim Uniform(0,1), \\ \alpha_0, \alpha_1, \alpha_2, a, b \sim N(0, 100^2), \\ 1/\sigma^2, 1/\sigma_s^2 \sim Gamma(0.01, 0.01), \\ \gamma \sim N(1,5). \end{cases}$$

Posterior inference will be drawn for the dose-response at time t of clinical interest and the 95% credible intervals will also be plotted.

4.6. Safety Analyses

Unless specified otherwise, safety assessments will be guided by an estimand comparing safety of LY3437943 doses with placebo irrespective of adherence to study drug. Thus, safety analyses will be conducted using the safety analysis set.

4.6.1. Extent of Exposure

Listing of exposure to LY3437943 and placebo will be provided by treatment group using data from the SS. Summary of duration of follow-up (defined as time in days from date of randomization to the date of the last study visit) and/or duration on study treatment (defined as time in days from date of first dose of study treatment to date of last dose of study treatment plus 7 days) will be provided by treatment group using data from SS, in the following period:

- 48 weeks plus safety follow-up (Visit 801) for all randomized participants

For the summary of duration on study treatment, the frequency and percentage of participants falling into the following range will be summarized by planned treatment group as well:

- greater than 0
- 4 weeks or longer
- 8 weeks or longer
- 12 weeks or longer
- 16 weeks or longer
- 20 weeks or longer
- 24 weeks or longer
- 28 weeks or longer
- 32 weeks or longer
- 36 weeks or longer
- 40 weeks or longer
- 44 weeks or longer, and
- 48 weeks or longer.

In addition, the frequency and percentages of participants falling into the following study treatment exposure ranges may be summarized by planned treatment group:

- 0 weeks,
- longer than 0 weeks to less than 4 weeks
- 4 weeks or longer to less than 8 weeks
- 8 weeks or longer to less than 16 weeks
- 16 weeks or longer to less than 20 weeks
- 20 weeks or longer to less than 24 weeks
- 24 weeks or longer to less than 28 weeks
- 28 weeks or longer to less than 32 weeks
- 32 weeks or longer to less than 36 weeks
- 36 weeks or longer to less than 40 weeks
- 40 weeks or longer to less than 44 weeks, and
- 44 weeks or longer to less than 48 weeks.

No p-values will be reported in these summaries as they are intended to describe the study populations rather than test hypotheses about them.

4.6.2. Adverse Events

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

For events occurring on the day of first taking study medication, the CRF-collected information (eg, treatment emergent flag, start time of study treatment and event) will be used to determine whether the event was pre versus posttreatment if available. If the relevant information is not available, then the events will be counted as posttreatment.

The counts and percentages of participants with TEAEs will be summarized by treatment using MedDRA PT nested within SOC. Statistical comparisons will be applied at both the SOC and PT levels. Events will be ordered by decreasing frequency within SOC. The SOC will be in alphabetical order. For events that are sex-specific, the denominator and computation of the percentage will include only participants from the given sex.

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

The counts and percentages of patients with TEAEs by maximum severity will be summarized by treatment using MedDRA PT. For each participant and TEAE, the maximum severity for the MedDRA PT is the maximum postbaseline severity observed from all associated LLTs mapping to the MedDRA PT. The maximum severity will be determined based on the nonmissing

severities. If all severities are missing for the defined postbaseline period of interest, it will show as missing in the table.

4.6.3. Patient Narratives

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

- death
- SAE, or
- permanent discontinuation of study treatment due to AEs.

Patient narratives (patient level data and summary paragraph) will be provided for participants in the randomized population with at least 1 notable event.

4.6.4. Vital Signs

In the case that multiple records of an individual vital sign are collected at the same visit, they will be averaged prior to being used for data summaries and analyses.

Descriptive summaries by treatment and by nominal visit will be provided for baseline and postbaseline values as well as change from baseline values.

Treatment differences in mean change will be analyzed using the MMRM model as described in Section 4.1 with all available data in the safety analysis set.

Counts and percentages of participants with treatment-emergent abnormal sitting systolic blood pressure, sitting diastolic, and pulse will be presented by treatment for participants who have both baseline and at least 1 postbaseline result. A treatment-emergent high result is defined as a change from a value less than or equal to the high limit at baseline to a value greater than the high limit at any time that meets the specified change criteria during the postbaseline period. A treatment-emergent low result is defined as a change from a value greater than or equal to the low limit at baseline to a value less than the low limit at any time that meets the specified change criteria during the postbaseline period. To assess decreases, change from the minimum value during the baseline period to the minimum value during the postbaseline period will be used. To assess increases, changes from the maximum value during the baseline period to the maximum value during the postbaseline period will be used. Both planned and unplanned measurements will be included in the analysis. The criteria for identifying patients with treatment-emergent vital sign abnormalities are stated in Table GZBF.4.6.

Table GZBF.4.6. Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurements

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

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

The treatment-emergent high systolic blood pressure will also include the criterion (≥140 and increase from baseline ≥20). The data summaries and analyses specified in this section will also apply to the measurements from ABPM where the same criteria applied to pulse rate will be applied to the heart rate.

In addition, following analyses will be conducted by treatment:

- counts and percentages of participants who had resting heart rate changes from baseline at 2 consecutive visits of more than 10 bpm and/or 20 bpm
- counts and percentages of participants who had at least 1 resting heart rate exceeding 100 bpm, and
- counts and percentages of participants who had at least 1 resting heart rate exceeding 100 bpm occurring at 2 consecutive study visits.

4.6.5. Electrocardiograms

Summary statistics by treatment and by nominal visit will be provided for ECG parameters (heart rate, PR, QRS, RR, and QT corrected using QTcF). When the QRS is prolonged (eg, a complete bundle branch block), QTcF should be used to assess ventricular repolarization. Thus, for a particular ECG, the following will be set to missing (for analysis purposes) when QRS is 120 msec or longer: QT and QTcF. Change from baseline to postbaseline values for ECG parameters will be summarized for patients who have both a baseline and at least 1 postbaseline result. Only planned measurements will be included in the mean change analyses. The criteria for identifying participants with treatment-emergent quantitative ECG abnormalities is based on Table GZBF.4.7. The counts and percentages of participants who meet following criteria at any time during the entire study period (including the off drug follow up time period) will be summarized by treatment group:

- treatment-emergent ECG abnormalities as listed in Table GZBF.4.7
- PR greater than or equal to 220 msec with 0% through 25% and higher than 25% increase from baseline
- QTcF greater than 500 msec, and
- treatment-emergent increase from baseline in QTcF interval of greater than 30 msec, 60 msec, or 75 msec.

Treatment-emergent qualitative ECG abnormalities are defined as qualitative abnormalities that first occurred after baseline. A listing of abnormal qualitative ECGs will be created (see).

Table GZBF.4.7. Selected Categorical Limits for ECG Data

	Low		High	
Parameter	**Males**	**Females**	**Males**	**Females**
Heart Rate (bpm)	<50 and decrease ≥15	<50 and decrease ≥15	>100 and increase ≥15	>100 and increase ≥15
PR Interval (msec)	<120	<120	≥220	≥220
QRS Interval (msec)	<60	<60	≥120	≥120
RR Interval (msec)	<300	<300	>1714	>1714
QTcF (msec)	<330	<340	>450	>470

Abbreviations: bpm = beats per minute; ECG = electrocardiogram; QTcF = Fridericia's corrected QT interval.

4.6.6. Clinical Laboratory Evaluation

All laboratory data will be reported in SI units. Selected laboratory measures will also be reported using CN units. Limits from the performing lab will be used to define low (L) and high (H). Descriptive summaries by treatment and by nominal visit will be provided for the baseline and postbaseline values as well as the change from baseline values.

Observed and change from baseline values for each visit may be displayed in plots for participants who have both a baseline and at least 1 postbaseline planned measurement. Baseline will be the last nonmissing observation prior to taking first study drug. Unplanned measurements will be excluded from plots.

A shift table will be provided including unplanned measurements. The shift table will include the number and percentage of participants within each baseline category (low, normal, high, or missing) versus each postbaseline category (low, normal, high, or missing) by treatment. The proportion of participants shifted will be compared between treatments using Fisher's exact test.

For qualitative laboratory analytes, the number and percentage of participants with normal and abnormal values will be summarized by treatment.

A listing of abnormal findings will be created for laboratory analyte measurements, including qualitative measures. The listing will include participant identification, treatment group, laboratory collection date, study day, analyte name, and analyte finding.

The MMRM model as described in Section 4.3.2 or ANHECOVA (if MMRM model is not applicable) as described in Section 4.3.3 will be used for the analysis during the treatment period for the continuous measurements for selected lab tests. The ANHECOVA, if used in this case, will use EAS with multiple imputation performed by applying only hypothetical strategy to ICE.

4.6.7. Additional Safety Assessments

4.6.7.1. Exocrine Pancreas Safety

4.6.7.1.1. Pancreatic Enzyme

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

The counts and percentages of participants with maximum postbaseline pancreatic enzyme value exceeding the following thresholds will be provided by baseline pancreatic enzyme value (less than or equal to 1× ULN, greater than 1× ULN), and treatment: less than or equal to 1× ULN, (less than 1 to greater than or equal to 3) × ULN, (greater than 3 to less than or equal to 5) × ULN, (greater than 5 to less than or equal to 10) × ULN, greater than 10× ULN. Missing will be considered as a separate group when calculating the counts and percentages.

An MMRM analysis as described in Section 4.3.2 on SS will be used to analyze each pancreatic enzyme with log transformed (postbaseline measure/baseline measure) response variables.

4.6.7.1.2. Pancreatic Events

Summaries of adjudicated and investigator-reported pancreatic events will be provided by treatment. Detailed searching criteria can be found in Section 7.6 (Appendix 6).

4.6.7.2. Gastrointestinal Safety

4.6.7.2.1. Nausea, Vomiting, and Diarrhea

Summaries and analyses for incidence and severity of nausea, vomiting, diarrhea, and 3 events combined, will be provided by each treatment group.

Summary of the prevalence over time for nausea, vomiting, and diarrhea will also be presented.

Time to the onset of nausea, vomiting, and diarrhea will be plotted. Prevalence and incidence of treatment-emergent nausea, vomiting, and diarrhea will also be plotted.

4.6.7.2.2. Severe Gastrointestinal Events

Severe GI AEs (GI SOC) will be captured with AE-CRF form and serious cases will be captured with the SAE form. The PTs in the most updated version of GI SOC MedDRA will be used to identify GI AEs, and only the PTs with serious/severe cases will be considered as AESIs.

The counts and percentages of participants with severe GI events will be summarized by treatment.

4.6.7.3. Hepatic Safety

4.6.7.3.1. Hepatobiliary Disorders

Hepatobiliary disorders will be considered as AESIs. The counts and percentages of participants with treatment-emergent potentially drug-related hepatic disorders will be summarized by treatment using the MedDRA PTs. Detailed searching criteria can be found in Section 7.6 (Appendix 6).

4.6.7.3.2. Liver Enzymes

Analyses for laboratory analyte measurements are described in Section 4.6.6. This section describes additional analyses of liver enzymes.

The counts and percentages of participants with the following elevations in hepatic laboratory tests at any time during the entire study including follow up period will be summarized between treatment groups:

- ALT and AST
  The counts and percentages of participants whose maximum postbaseline results are less than or equal to 1× ULN, (greater than 1 to less than or equal to 3) × ULN, (less than 3 to greater than or equal to 5) × ULN, (greater than 5) × ULN or missing will be summarized by treatment group by maximum baseline result in less than or equal to 1.5× ULN], greater than 1.5× ULN or missing.
- TBL and direct bilirubin
  The counts and percentages of participants whose maximum post-baseline results are less than or equal to 1× ULN, (greater than 1 to less than or equal to 2) × ULN, greater than 2× ULN or missing will be summarized by treatment group by maximum baseline result in greater than or equal to 1.5× ULN, less than 1.5 × ULN or missing.
- ALP
  The counts and percentages of participants whose maximum post-baseline results are less than or equal to 1× ULN, (greater than 1 to less than or equal to 2) × ULN, (greater than 2 to less than or equal to 3) × ULN, greater than 3× ULN or missing will be summarized by treatment group by maximum baseline result in greater than or equal to 1.5× ULN, less than 1.5 × ULN or missing.

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

Two plots will be provided as follows:

- Hepatocellular DILI Screening Plot (TBL versus ALT): Each patient is plotted (ie, scatterplot) based on their maximum postbaseline TBL (y-axis) versus ALT, regardless of the time between the 2 maximum values. Dashed lines represent TBL and transaminase cut-offs of 2× ULN and 3× ULN (default) respectively. A potential Hy's Law case is circled and defined as having a maximum postbaseline TBL greater than or equal to 2× ULN within 30 days after maximum postbaseline ALT or AST greater than or equal to 3× ULN, without findings of cholestasis (defined as ALP less than 2× ULN). Include all scheduled and unscheduled laboratory test values.
- Hepatocellular DILI Screening Plot (TBL versus AST): Each patient is plotted (ie, scatterplot) based on their maximum postbaseline TBL (y-axis) versus AST, regardless of the time between the 2 maximum values. Dashed lines represent TBL and transaminase cut-offs of 2× ULN and 3× ULN (default) respectively. A potential Hy's Law case is circled and defined as having a maximum postbaseline TBL greater than or equal to 2× ULN within 30 days after maximum postbaseline ALT or AST greater than or equal to

3× ULN, without findings of cholestasis (defined as ALP less than 2× ULN). Include all scheduled and unscheduled laboratory test values.

The counts and percentages of participants in each quadrant of the respective plots may be provided by treatment group of LY3437943 and placebo, as shown in Table GZBF.4.8 and Table GZBF.4.9 if data warranted.

Table GZBF.4.8. Summary of Participants with Potential Hepatocellular DILI in LY3437943 Group Versus Placebo

Quadrant	LY3437943 (N=XXX) n (%)	Placebo (N=XXX) n (%)
Potential Hy's Law (Right upper)		
Cholestasis (Left upper)		
Temple's corollary (Right lower)		
Total		

Abbreviations: DILI = drug-induced liver injury; N = number of patients in the analysis population; n = number of patients in the specified category.

Table GZBF.4.9. Summary of Participants with Potential Cholestatic DILI in LY3437943 Group Versus Placebo

Quadrant	LY3437943 (N=XXX) n (%)	Placebo (N=XXX) n (%)
TBL ≥2× ULN and ALP ≥2× ULN (Right upper)		
TBL ≥2× ULN and ALP <2× ULN (Left upper)		
TBL <2× ULN and ALP <2× ULN (Left lower)		
TBL <2× ULN and ALP ≥2× ULN (Right lower)		
Total		

Abbreviations: ALP = alkaline phosphatase; DILI = drug-induced liver injury; N = number of patients in the analysis population; n = number of patients in the specified category; TBL = total bilirubin; ULN = upper limit of normal.

4.6.7.4. Hypoglycemia

Per the study protocol, investigators should use the following definitions and criteria when diagnosing and categorizing an episode considered to be related to hypoglycemia (the blood glucose values in this section refer to values determined by a laboratory or International Federation of Clinical Chemistry and Laboratory Medicine blood-equivalent glucose meters and strips) in accordance with the 2020 American Diabetes Association position statement on glycemic targets (ADA 2020) as below. Level 2 and Level 3 hypoglycemia events are considered as safety topics of special interest.

- **Level 1 Hypoglycemia (Level 1):**
  Glucose less than 70 mg/dL (3.9 mmol/L) and glucose greater than or equal to 54 mg/dL (3.0 mmol/L)

- **Level 2 Hypoglycemia (Level 2):**
  Glucose less than 54 mg/dL (3.0 mmol/L)
- **Severe Hypoglycemia (Level 3):**
  Severe hypoglycemia (in adults): a severe event characterized by altered mental and/or physical status requiring assistance for treatment of hypoglycemia. For example, participants had altered mental status and could not assist in their own care or were semiconscious or unconscious, or experienced coma with or without seizures, and the assistance of another person was needed to actively administer carbohydrate, glucagon, or other resuscitative actions. Glucose measurements may not be available during such an event, but neurological recovery attributable to the restoration of glucose concentration to normal is considered sufficient evidence that the event was induced by a low glucose concentration.
    - The determination of a hypoglycemic event as an episode of severe hypoglycemia, as defined above, is made by the investigator based on the medical need of the participant to have required assistance and is not predicated on the report of a participant simply having received assistance.
    - If a hypoglycemic event meets the criteria of severe hypoglycemia, the investigator must record the event as serious on the AE CRF and report it to Lilly as an SAE.
- **Other hypoglycemia categories:**
  Nocturnal hypoglycemia is a hypoglycemia event (including severe hypoglycemia) that occurs at night and presumably during sleep.

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

Both the incidence (percent of patients experiencing 1 episode or more) and the rate (episodes/patient/year) of level 2 or level 3 hypoglycemia, and level 1 hypoglycemia will be reported by treatment group. For these analyses, only hypoglycemia events prior to permanent discontinuation of study drug will be included. A listing of all events of severe hypoglycemia may be provided, if deemed necessary. This listing will provide treatment allocation, clinical characteristics of the hypoglycemic event, and concomitant medications.

Summary of level 1 hypoglycemia will be provided for the data in SS and the data in SS excluding hypoglycemic events occurring after initiation of a new antihyperglycemic therapy.

4.6.7.5. Immunogenicity

Treatment-emergent anti-drug antibodies are defined as those with a titer 2-fold (1 dilution) greater than the minimum required dilution if no ADAs were detected at baseline (treatment-induced ADA) or those with a 4-fold (2 dilutions) increase in titer compared to baseline if ADAs were detected at baseline (treatment-boosted ADA). A patient is evaluable for TE ADA if the patient has a nonmissing baseline ADA result, and at least 1 nonmissing postbaseline ADA result.

Listings of patients who are not TE ADA evaluable, patients with at least 1 test having detected LY3437943 ADAs, and patients having LY3437943 ADAs present or TEAE: hypersensitivity reactions or injection site reactions will be provided.

The frequency and percentage of patients with preexisting ADA, with TE ADA, with cross-reactive antibodies and with neutralizing antibodies will be tabulated by dose (if data warrant), where proportions are relative to the number of patients who are TE ADA evaluable. The frequency and percentage of patients with hypersensitivity and injection site reaction TEAEs by TE ADA status will be tabulated if data warrant.

4.6.7.6. Hypersensitivity Events

Hypersensitivity reactions and related information reported in electronic case report form (eCRF) will be listed and summarized by treatment.

Two main analyses are performed:

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

Summaries of all potential hypersensitivity reactions will be generated by PT with decreasing frequency by treatment. The AE database will be searched using predefined SMQs to identify events consistent with hypersensitivity events. Severe/serious hypersensitivity events identified by predefined SMQ search will be considered as AESIs.

4.6.7.7. Injection Site Reactions

Injection site reactions, incidence, and related information reported in eCRF will be summarized by treatment. Information to be summarized includes the timing of the reaction relative to study drug administration, and characteristics of the injection site reaction: erythema, induration, pain, pruritus, and edema.

Additionally, potential injection site reactions will be searched by predefined MedDRA HLTs of injection site reactions, administration site reactions, and infusion related reactions. Detailed searching criteria for injection site reaction events can be found in Section 7.6 (Appendix 6). The PT will be used for summary by treatment within each HLT category. Only the severe/serious injection site reactions will be considered as AESI.

4.6.7.8. Renal Safety

Laboratory measures related to renal safety will be analyzed as specified for laboratory measurements in Section 4.6.6.

Two shift tables examining renal function will be created. A min-to-min shift table of eGFR estimated by the CKD-EPI equation with unit mL/min/1.73 m^2, using categories (less than 30, greater than or equal to 30 to less than 45, greater than or equal to 45 to less than 60, greater than

or equal to 60 to less than 90, and greater than or equal to 90 mL/min/1.73 m^2). A max-to-max shift table of UACR, using the categories UACR less than 30 mg/g, greater than or equal to 30 mg/g and less than or equal to 300 mg/g, UACR greater than 300 mg/g (respectively, these represent normal, microalbuminuria, and macroalbuminuria).

Mixed model repeated measure analyses for eGFR and UACR will be provided. Log transformation will be performed for UACR.

4.6.7.8.1. Acute Renal Events

Because severe GI events may lead to dehydration, which could cause a deterioration in renal function including acute renal failure, dehydration events will be analyzed. Acute renal events associated with chronic renal failure exacerbation will also be captured.

Acute renal events will be considered as AESI.

The counts and percentages of participants with acute renal events will be summarized by treatment by using the MedDRA PTs contained in any of the following SMQs:

- Acute renal failure: narrow terms in Acute renal failure SMQ (20000003), and
- Chronic kidney disease: narrow terms in Chronic kidney disease SMQ (20000213).

In addition, a listing of participants with treatment-emergent acute renal events may be provided, if deemed necessary.

4.6.7.8.2. Dehydration

Dehydration events will be captured in the narrow terms in Dehydration SMQ (20000232).

A listing of participants with treatment-emergent dehydration events will be provided.

4.6.7.9. Thyroid Safety Monitoring

4.6.7.9.1. Calcitonin

Observed calcitonin data (a thyroid-specific laboratory assessment) will be summarized by treatment and nominal visit.

The counts and percentages of participants with a maximum postbaseline calcitonin value in the following thresholds will be provided by treatment and baseline calcitonin value (20 ng/L or lower, higher than 20 ng/L to 35 ng/L or lower, higher than 35 ng/L). Postbaseline: 20 ng/L or lower, higher than 20 ng/L to 35 ng/L or lower, higher than 35 ng/L to 50 ng/L or lower, higher than 50 ng/L to 100 ng/L or lower, and higher than 100 ng/L.

4.6.7.9.2. C-Cell Hyperplasia and Thyroid Malignancies

Thyroid malignancies and C-cell hyperplasia will be considered as AESI. Treatment-emergent thyroid malignancies and C-cell hyperplasia will be identified using Malignancies SMQ (20000090), MedDRA HLT for Thyroid neoplasms and PT for Thyroid C-cell hyperplasia.

The counts and percentages of participants with treatment-emergent thyroid C-cell hyperplasia and malignancies will be summarized by treatment and PT ordered with decreasing frequency. In addition, a listing of participants with treatment-emergent thyroid C-cell hyperplasia and neoplasms may be provided if deemed necessary.

4.6.7.10. Major Adverse Cardiovascular Events

Major adverse cardiovascular events reported by investigators are adjudicated by an independent CEC in a blinded fashion.

The CV AEs to be adjudicated include deaths due to CV cause, myocardial infarction, hospitalization for unstable angina, hospitalization for heart failure, coronary interventions (such as coronary artery bypass graft or percutaneous coronary intervention); and cerebrovascular events, including cerebrovascular accident (stroke) and transient ischemic attack.

Only adjudicated MACE will be considered as AESI. The counts and percentages of participants with adjudicated MACE may be summarized by treatment.

In addition, MACE reported by investigator may also be summarized although a MACE reported by investigator is not considered as AESI.

A listing of participants reporting MACE events, either reported by investigator or identified by the CEC, will be provided. The listing will include treatment, participants identification including the site number, date of event, type of event as reported by the investigator, type of event as adjudicated by the CEC, time from first dose of study drug to the event, and time from last dose to the event (if participant has discontinued study drug prior to the event).

4.6.7.11. Treatment-Emergent Supraventricular Arrhythmias and Cardiac Conduction Disorders

Treatment-emergent supraventricular arrhythmias and cardiac conduction disorders will be considered as AESI. The CV events will include clinically relevant rhythm and conduction disorders.

The treatment-emergent supraventricular arrhythmias and cardiac conduction disorders events will be included using the MedDRA PTs. Detailed searching criteria can be found in Section 7.6 (Appendix 6).

The counts and percentages of participants with treatment emergent supraventricular arrhythmias and cardiac conduction disorders will be summarized by treatment and PT nested within SMQ. The PT will be ordered with decreasing frequency within SMQ. A listing of participants with treatment-emergent supraventricular arrhythmias and cardiac conduction disorders may be provided if deemed necessary.

4.6.7.12. Major Depressive Disorder/Suicidal Ideation

The major depressive disorder/suicidal ideation or behavior will be captured as AESI. Adverse events will be searched using MedDRA PT terms. Detailed searching criteria can be found in Section 7.6 (Appendix 6).

The counts and percentages of participants with TEAEs will be summarized by treatment group using MedDRA PT nested within SMQ. Events will be ordered by decreasing frequency in the total LY3437943 group nested within SMQ. A listing of participants with major depressive disorder/suicidal ideation or behavior may be provided if deemed necessary.

Additionally, suicidal ideation and behavior, and depression will be assessed by the investigator via spontaneously reported AEs and through the use of the Columbia-Suicide Severity Rating Scale (C-SSRS) and the PHQ-9.

4.6.7.12.1. Patient Health Questionnaire

Total scores for the PHQ-9 range from 0 to 27 with total scores categorized as

- none (not depressed): 0 through 4
- mild: 5 through 9
- moderate: 10 through 14
- moderately severe: 15 through 19, and
- severe: 20 through 27.

Shift tables will be provided showing the counts and percentages of participants within each baseline category (maximum value) versus each postbaseline category (maximum value) by treatment.

Additionally, the following 3 outcomes of interest will be compared between treatments (based on the maximum value during baseline and postbaseline):

- any increase in depression category (ie, worsening of depression): includes participants in the none, mild, moderate, or moderately severe category during baseline and with at least 1 postbaseline measurement
- increase from No or Mild Depression to Moderate, Moderately Severe or Severe Depression: includes participants in the none or mild depression category during baseline and with at least 1 postbaseline measurement; and
- increase from Mild or Moderate Depression to Moderately Severe or Severe Depression: includes participants in the mild or moderate depression category during baseline and with at least 1 postbaseline measurement.

4.6.7.12.2. Suicidal Ideation and Behavior Solicited Through C-SSRS

Suicide-related thoughts and behaviors occurring during the entire study period, based on the C-SSRS, will be summarized by treatment group. In particular, for each of the following suicide-related events, the counts and percentages of participants with the event will be summarized by treatment group:

- died by suicide
- nonfatal suicide attempt
- interrupted attempt
- aborted attempt
- preparatory acts or behavior
- active suicidal ideation with specific plan and intent
- active suicidal ideation with some intent to act without specific plan
- active suicidal ideation with any methods (no plan) without intent to act
- nonspecific active suicidal thoughts
- wish to be dead, and
- nonsuicidal, self-injurious behavior.

In addition, the counts and percentages of participants who experienced at least 1 of the composite measures will be presented. The participants with at least 1 postbaseline C-SSRS assessment are included. The composite measure is determined at each assessment by the "yes" or "no" responses in C-SSRS categories by the study participant:

- Category 1 – Wish to be Dead
- Category 2 – Non-specific Active Suicidal Thoughts
- Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
- Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
- Category 5 – Active Suicidal Ideation with Specific Plan and Intent
- Category 6 – Preparatory Acts or Behavior
- Category 7 – Aborted Attempt
- Category 8 – Interrupted Attempt
- Category 9 – Actual Attempt (non-fatal), and
- Category 10 – Completed Suicide.

Composite endpoints of suicidal ideation and suicidal behavior based on the above categories are defined below:

- Suicidal ideation: A "yes" answer at any time during study to any 1 of the 5 suicidal ideation questions (Categories 1 through 5) on the C-SSRS.
- Suicidal behavior: A "yes" answer at any time during study to any 1 of the 5 suicidal behavior questions (Categories 6 through 10) on the C-SSRS.
- Suicidal ideation or behavior: A "yes" answer at any time during study to any 1 of the 10 suicidal ideation and behavior questions (Categories 1 through 10) on the C-SSRS.

A listing contains data for each participant with suicidal ideation, suicidal behavior, or non-suicidal self-injurious behavior during the study by treatment and visit. Data from all visits are displayed, regardless of a "yes" or "no" answer, for patients with any "yes" answer for suicidal ideation, suicidal behavior, or self-injurious behavior without suicidal intent.

4.6.7.13. CCI

4.6.7.14. Drug Abuse, Dependence and Withdrawal

Subjects with drug abuse, dependence and withdrawal will be listed. Related events can be identified by searching the MedDRA PTs contained in any of the following SMQs:

- Broad and narrow terms in drug abuse and dependence SMQ (20000101) and
- Broad and narrow terms in drug withdrawal SMQ (20000102).

4.7. Other Analyses

4.7.1. Health Outcomes

The patient-reported outcome questionnaires will be analyzed using the EAS, unless specified otherwise.

Item-level missingness is dealt with as per the instrument developers' instruction.

Additional psychometric analyses may be performed by Global Patient Outcomes Real World Evidence at Lilly and documented in a separate analysis plan.

4.7.1.1. Patient Global Impression of Status for Physical Activity

The counts and percent of participants for PGIS response categories at each time point will be summarized by nominal visit and by treatment. A shift table from baseline to postbaseline of 5 PGIS response categories will be created at each postbaseline visit.

4.7.1.2. Short-Form-36 Health Survey Version 2, Acute Form

Per copyright owner, QualityMetric Health Outcomes Scoring (PRO_CoRe V2.0) will be used to derive the following domain and component scores:

- Mental Component Summary Score (MCS)
- Physical Component Summary Score (PCS)
- Physical Functioning domain (PF)
- Role-Physical domain (RP)
- Bodily Pain domain (BP)
- General Health domain (GH)
- Vitality domain (VT)
- Social Functioning domain (SF)
- Role-Emotional domain (RE), and
- Mental Health domain (MH).

For each above parameter, the raw scores will be transformed into the domain scores (t-scores) and the following analyses for the actual value and change from baseline value will be conducted on EAS:

- Descriptive summaries by treatment group, and
- analysis from MMRM (as described in Section 4.3.2)/ANHECOVA model (as described in Section 4.3.3) depending on whether there is more than 1 postbaseline response variables. The multiple imputation will use only hypothetical strategy to deal with ICE if ANHECOVA model is used.

4.7.1.3. Eating Inventory

The following domain and component scores related to eating behavior will be derived based on the 51-item Eating Inventory questionnaire:

- dietary restraint (21 items)
- disinhibition (16 items), and
- perceived hunger (14 items).

For each above parameter, the raw scores will be transformed into the domain scores and the following analyses for the actual value and change from baseline value will be conducted:

- descriptive summaries by treatment group, and
- analysis from MMRM (as described in Section 4.3.2)/ANHECOVA model (as described in Section 4.3.3) depending on whether there is more than 1 postbaseline response variables. The multiple imputation will use only hypothetical strategy to deal with ICE if ANHECOVA model is used.

Subjects who do not speak English as their native language may have missing baseline data and will be excluded from this analysis.

4.7.2. Subgroup analyses

Subgroup analyses of the primary endpoint will be made to assess consistency of the intervention effect across the following subgroups:

- age group: younger than 65 versus 65 years or older
- race
- sex: female versus male
- ethnicity
- BMI (kg/m^2) group: less than 30 versus greater than or equal to 30 and less than 35 versus greater than or equal to 35 and less than 40 versus greater than or equal to 40, and
- baseline BMI (the minimum or greater to less than the first quartile, the first quartile or greater to less than the median, the median or greater to less than the third quartile, the third quartile or greater to the maximum).

For each subgroup analysis, MMRM model as described in Section 4.3.2 will be used.

If the number of participants is too small (less than 10%) within a subgroup, then the subgroup categories may be redefined prior to unblinding the study. Additional subgroup analyses may also be performed.

4.8. Interim Analyses

An interim analysis for Study GZBF will be conducted when approximately 15% of the participants complete the Week 16 visit to evaluate the safety and tolerability profile of LY3437943. This safety interim analysis was planned to address potential tolerability concerns. It was planned after the approval of the protocol and thus was not documented in the protocol. A second interim analysis for Study GZBF will be conducted when approximately 100% of the participants complete the Week 16 visit. This interim analysis will evaluate both safety and efficacy profile of LY3437943 and may be used to determine the doses of future studies of LY3437943. A third interim analysis will be conducted when 100% of the participants complete the 36-week treatment period. This interim analysis will evaluate both safety and efficacy profile of LY3437943 and may be used to support end of phase 2 interactions with regulatory agencies. The second and third interim analyses were planned after the approval of the protocol and thus were not documented in the protocol.

An AC (assessment committee) will be formed to review the interim analyses in an unblinded manner. The details regarding endpoints of analysis and number of participants will be provided in the AC charter. Since the endpoints for interim analysis is a subset of the endpoints in Section 1.1, the analysis methods of these endpoints should be consistent with the ones described in this document.

Study team members who have potential contact with the sites will remain blinded throughout the study. Information that may unblind the study during the analyses will not be reported to study sites or blinded study team members before the study has been unblinded. Study sites will receive information about interim results only if deemed necessary for the safety of the participants.

5. Sample Size Determination

Approximately 300 participants will be randomly assigned to LY3437943 1 mg, LY3437943 4 mg (a), LY3437943 4 mg (b), LY3437943 8 mg (a), LY3437943 8 mg (b), LY3437943 12 mg, or placebo with a 2:1:1:1:1:2:2 ratio. Assuming a 20% dropout rate, approximately 48 participants will complete each LY3437943 dose and placebo. An upper limit of 60% enrollment of women will be used to ensure a sufficiently large sample of men.

Sample size selection is guided by the objective of establishing superiority of each LY3437943 dose to placebo relative to the percent change in body weight from baseline to Week 24. The evaluation of superiority to placebo will be conducted for each of the 4 LY3437943 doses at 2-sided significance level of 0.05 using 2-sample t-test. Assuming a SD of 10%, the LY3437943 group mean percent change in body weight at Week 24 from baseline compared to placebo is assumed to be -8%. The chosen sample size provides at least 97% power to establish superiority of LY3437943 1 mg, LY3437943 4 mg, LY3437943 8 mg, or LY3437943 12 mg compared to placebo. No adjustment for multiplicity will be performed.

In addition, this sample size provides approximately 83% power to show that at least 1 LY3437943 dose has superior effects of reducing waist circumference compared with placebo at Week 24. This assumes a reduction of 5.8 cm in waist circumference compared with placebo at Week 24 and a SD of 9.6 cm.

6. Novel Coronavirus Disease 2019 Impact

The following additional statistical analyses may be performed at the primary database lock and final database lock to assess the impact of COVID-19 pandemic for all randomized participants if the data warrants:

- listing of all randomized participants who discontinue study due to COVID-19 pandemic
- listing of all study disruptions related to COVID-19 pandemic
- listing of AEs or deaths related to COVID-19 pandemic, and
- listing of important protocol deviations due to COVID-19 pandemic.

In case there is a larger impact of COVID-19 on the study, due to a shut-down or any other reason, more details for additional analyses may be provided.

For the primary endpoints and key secondary endpoints, missing data due to COVID-19 will be handled as described in Section 4.3.3.

7. Supporting Documentation

7.1. Appendix 1: Demographic and Baseline Characteristics

A listing of participant demographics for all randomized participants will be provided. All demographic and baseline clinical characteristics will be summarized by treatment groups and dose escalation subgroups for all randomized participants.

Baseline demographic and clinical characteristics of special interest include, but are not limited to, age (years), sex (female, male), race, ethnicity, height (cm), weight (kg), BMI (kg/m^2), waist circumference (cm), age group (younger than 65 years, 65 years or older), BMI group (less than 30, greater than or equal to 30 and less than 35, greater than or equal to 35 and less than 40, greater than or equal to 40 kg/m^2).

7.2. Appendix 2: Historical Illnesses and Preexisting Conditions

The count and percentages of participants with historical illnesses and preexisting conditions will be summarized by treatment groups and dose escalation subgroups using the MedDRA PTs nested within SOC. The SOC will be in alphabetical order. Conditions (ie, PTs) will be ordered by decreasing frequency within SOC. This will be summarized for all randomized participants.

7.3. Appendix 3: Concomitant Medications

Concomitant medications will be summarized by treatment group. The percentages of participants who took concomitant medication will be summarized by treatment using PTs nested within ATC Level 3 codes. The concomitant medications will be ordered by decreasing frequency within each ATC level.

Concomitant medication will be summarized by PTs by treatment groups and dose escalation subgroups by decreasing frequency for the safety analysis set.

Additionally, medications of interest (as defined below) will be summarized by treatment groups and dose escalation subgroups for the safety analysis set.

Concomitant medications of interest include the following:

- baseline antihypertensive therapy, by type/class
- baseline lipid lowering therapy, by type/class
- changes to baseline medication in post randomization (in term of type/class and dose:
    - antihypertensive therapy, and
    - lipid lowering therapy
- utilization after randomization of:
    - antihyperglycemic medication for the treatment of diabetes for participants who develop T2DM during the study.

7.4. Appendix 4: Treatment Compliance

Listing and summary of prematurely discontinuing study treatment (including discontinuation reason) and discontinuing study will be provided by treatment groups and dose escalation subgroups. Kaplan-Meier plots of time to premature study treatment discontinuation and time to

premature study treatment discontinuation due to AE will be provided at Weeks 24 and 48 based on all randomized population. Kaplan-Meier plots of time to premature study discontinuation will be provided at Weeks 24 and 48 based on all randomized population. Time-to-event analyses of premature study treatment discontinuation, study treatment discontinuation due to AE, and study discontinuation may be conducted at Weeks 24 and 48.

If data warrants, the counts and percentages of participants who follow the planned dose escalation scheme, have dose interruption, or have dose de-escalation will be summarized for each treatment group and dose escalation subgroups. Listing of overdose of LY3437943 and placebo will be provided.

Per protocol, treatment compliance will be assessed every 4 weeks, or every 12 weeks after Visit 11, at the time of visits to the study site. Treatment compliance will be defined as taking at least 75% of the scheduled LY3437943 or placebo doses. Compliance at each treatment compliance assessment visit mentioned above and over the corresponding study period will be calculated using the number of doses administered (regardless of the actual dose in unit or mL administered) divided by the total number of doses expected to be administered ×100 at the specific visit or over the corresponding study period, respectively. Treatment compliance will be summarized descriptively at each treatment compliance assessment visit and over the corresponding study period by treatment groups and dose escalation subgroups using the full analysis set.

7.5. 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 groups and dose escalation subgroups will be provided at the end of 48 weeks treatment (for all randomized participants).

7.6. Appendix 6: Searching Criteria for Additional Safety Assessments

Pancreatitis Events

Determination of investigator-reported events will be through the “Acute pancreatitis” Standardized MedDRA Query (20000022, narrow scope) and a “Chronic pancreatitis” PT search of the AE database, while adjudication-confirmed pancreatitis are found from adjudication forms.

Treatment-Emergent Hepatobiliary Disorders

Treatment-emergent potentially drug-related hepatic disorders will be summarized by treatment using the MedDRA PTs contained in any of the following SMQs:

- Broad and narrow terms in the Liver related investigations, signs and symptoms SMQ (20000008)
- Broad and narrow terms in the Cholestasis and jaundice of hepatic origin SMQ (20000009)
- Broad and narrow terms in the Hepatitis non-infections SMQ (20000010)
- Broad and narrow terms in the Hepatic failure, fibrosis and cirrhosis and other liver damage SMQ (20000013)
- Narrow terms in the Liver-related coagulation and bleeding disturbances SMQ (20000015)
- Narrow PTs in Gallbladder related disorders SMQ (20000124)

- Narrow PTs in Biliary tract disorders SMQ (20000125); and
- Narrow PTs in Gallstone related disorders SMQ (20000127).

Injection Site Reactions

Treatment-emergent injection site reaction will be summarized by treatment using the MedDRA PT in any of the following:

- MedDRA HLT of Injection site reaction
- HLT of Administration site reactions NEC, and
- HLT of Infusion Site Reactions.

Supraventricular Arrhythmias and Cardiac Conduction Disorders

Treatment-emergent supraventricular arrhythmias, arrhythmias and cardiac conduction disorders will be considered as an AESI. The CV events will include clinically relevant rhythm and conduction disorders. The treatment-emergent supraventricular arrhythmias and cardiac conduction disorders events will be included using the MedDRA PT contained in any of the following SMQs:

- Supraventricular Arrhythmias:
    - For symptoms: Arrhythmia related investigations, signs and symptoms SMQ (20000051), narrow and broad terms
    - For supraventricular arrhythmias:
        - Supraventricular tachyarrhythmia SMQ (20000057), broad and narrow terms
        - Tachyarrhythmia terms, nonspecific SMQ (20000164), narrow terms only; and
        - Ventricular tachyarrhythmia SMQ (20000058), narrow terms only.
- Cardiac Conduction Disorders
    - Conduction defects SMQ (20000056), narrow terms only; and
    - Cardiac conduction disorders HLT (10000032), all PTs.

Major Depressive Disorder/Suicidal Ideation

The major depressive disorder/suicidal ideation or behavior will be captured as AESI. Adverse events will be searched using MedDRA PT terms. The PTs from the Depression and suicide/self-injury SMQ as defined in MedDRA (SMQs: 20000037 [Suicide/self-injury] and 20000167 [Depression (excl suicide and self-injury)]) will be summarized.

7.7. Appendix 7: Nonalcoholic Fatty Liver Disease

This section is applicable to the participants who are enrolled in the NAFLD addendum.

Baseline clinical characteristics and demographic variables (including but not limited to liver fat content and other variables described in Section 7.1 (Appendix 1) will be summarized by treatment for the participants who are enrolled in the MRI addendum.

7.7.1. Primary Efficacy Analyses

The primary objective of NAFLD addendum is to assess the effect of 1, 4, 8, and 12 mg doses of LY3437943 compared with placebo from baseline (Visit 2 at Week -3) to Week 24 for relative liver fat change measured by magnetic resonance imaging proton density fat fraction in participants with NAFLD.

An ANHECOVA model as described in Section 4.3.3 will be used for EAS to analyze the primary endpoint using the baseline LFC as a covariate. The treatment difference and LSM of the difference estimated from the model will be provided and 2-sided 95% CIs for mean change in relative liver fat from baseline to Week 24 will be derived.

7.7.2. Secondary Efficacy Analyses

The following secondary objectives are considered in NAFLD addendum:

To assess the effect of 1, 4, 8, and 12 mg doses of LY3437943 compared with placebo from baseline (Visit 2 at Week -3) to Week 24 in participants with NAFLD for

- absolute liver fat change measured by MRI-PDFF.

To assess the effect of 1, 4, 8, and 12 mg doses of LY3437943 compared with placebo from baseline (Visit 2 at Week -3) to Week 48 in participants with NAFLD for

- relative liver fat changes measured by MRI-PDFF, and
- absolute liver fat changes measured by MRI-PDFF.

To assess the effect of 1, 4, 8, and 12 mg doses of LY3437943 compared with placebo from baseline (Visit 2 at Week -3) to Weeks 24 and 48 in participants with NAFLD for

- percentage of participants achieving a 30% or higher relative liver fat reduction.

For continuous endpoints, MMRM model as described in Section 4.3.2 or ANHECOVA model as described in Section 4.3.3 will be used depending on whether there are more than 1 postbaseline response variables. For binary endpoints, a logistic regression will be used and will include treatment group (dose escalation regimens as different levels), stratification factors as fixed effects, and the baseline LFC as a covariate. Multiple imputation will be used to impute the missing continuous-valued relative liver fat reduction before deriving the binary endpoint. Only hypothetical strategy to deal with ICEs will be used for these analyses.

7.7.3. Exploratory Efficacy Analyses

The following exploratory objectives are considered in NAFLD addendum:

To assess the effect of 1, 4, 8, and 12 mg doses of LY3437943 compared with placebo from baseline (Visit 2 at Week -3) to Weeks 24 and 48 in participants with NAFLD for the following MRI-derived measures:

- percentage of participants achieving a 50% or higher relative liver fat reduction
- percentage of participants achieving a 5% or higher absolute liver fat reduction
- percentage of participants achieving a liver fat content lower than 5%
- visceral adipose tissue volume
- abdominal SC adipose tissue volume

- lean thigh muscle volume
- muscle fat infiltration, and
- additional MRI-derived measures may be explored.

To assess the effect of 1, 4, 8, and 12 mg doses of LY3437943 compared with placebo from baseline (Visit 4 at Week 0) to Weeks 24 and 48 in participants with NAFLD on the following serum NASH biomarkers:

- K-18 (hepatocyte apoptosis biomarker)
- enhanced liver fibrosis panel (fibrosis biomarker)
- Pro-C3 (fibrosis biomarker, a fragment of the NH2-terminal propeptide of type III procollagen)
- NIS4 panel (blood-based diagnostic test to identify at-risk NASH)
- Ferritin (marker of iron stores frequently elevated in NASH), and
- FIB-4 index (marker of advanced fibrosis).

To assess the relationship between LY3437943 dose and/or concentration and liver fat with associated biomarkers.

For continuous endpoints other than body composition parameters (visceral adipose tissue, abdominal SC adipose tissue, lean thigh muscle volume, and muscle fat infiltration), MMRM model as described in Section 4.3.2 or ANHECOVA model as described in Section 4.3.3 with EAS will be used depending on whether there are more than 1 postbaseline response variables. For binary endpoints, a logistic regression will be used and will include treatment group (dose escalation regimens as different levels), stratification factors as fixed effects, and the baseline LFC as a covariate. Multiple imputation will be used to impute the missing continuous-valued relative liver fat reduction before deriving the binary endpoint. Missing values due to ICEs or other causes will be imputed from the observations in the EAS within the same treatment arm. For the body composition parameters (visceral adipose tissue, abdominal SC adipose tissue, lean thigh muscle volume, and muscle fat infiltration), use treatment policy estimand with all the data available in safety analysis set for MMRM model in Section 4.3.2. If using ANHECOVA model in Section 4.3.3, multiple imputation will be done by treatment group with all available data in safety analysis set.

8. References

[ADA] American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes – 2020. *Diabetes Care*. 2020;43(suppl 1):S66-S76. https://doi.org/10.2337/dc20-s006

Fu H, Manner D. Bayesian adaptive dose-finding studies with delayed responses. *J Biopharm Stat*. 2010;20(5):1055-1070. https://doi.org/10.1080/10543400903315740

[ICH] International Council for Harmonisation. Addendum on estimands and sensitivity analysis in clinical trials. To the guideline on statistical principles for clinical trials. E9(R1). November 20, 2019. Accessed April 07, 20202. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

Luo J, Ruberg SJ, Qu Y. Estimating the treatment effect for adherers using multiple imputation. *Pharm Stat*. 2021; In press. https://doi.org/10.1002/pst.2184

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*. 2002; In press. https://doi.org/10.1002/pst.2204.

Qu Y, Cui X, Bray R, Sun N. Efficient estimation of the efficacy and safety endpoints for clinical trials with preplanned dose titrations. Working paper; 2021

Qu Y, Liu Z, Fu H, et al. Modeling the impact of preplanned dose titration on delayed response. *J Biophar Stat*. 2019;29(2):287-305. https://doi.org/10.1080/10543406.2018.1535499

Ye T, Shao J, Yi Y, Zhao Q. Toward better practice of covariate adjustment in analyzing randomized clinical trials. 2009. Accessed March 30, 2022. https://arxiv.org/pdf/2009.11828.pdf

Signature Page for VV-CLIN-076021 v1.0

Approval	PPD PPD 15-Dec-2022 14:20:10 GMT+0000

Signature Page for VV-CLIN-076021 v1.0