I8F-MC-GPHR: Statistical Analysis Plan Version 2

A Randomized, Double-Blind, Placebo-Controlled Phase 2 Study Comparing the Efficacy and Safety of Tirzepatide versus Placebo in Patients with Nonalcoholic Steatohepatitis (NASH)

NCT04166773

Approval Date: 25-Jan-2024

Title Page

Protocol Title: A Randomized, Double-Blind, Placebo-Controlled Phase 2 Study Comparing the Efficacy and Safety of Tirzepatide versus Placebo in Patients with Nonalcoholic Steatohepatitis (SYNERGY-NASH)

Protocol Number: I8F-MC-GPHR

Compound Number: LY3298176

Short Title: Tirzepatide versus placebo in NASH

Acronym: SYNERGY-NASH

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana USA 46285

Regulatory Agency Identifier Number(s)

Registry	ID
IND	143191
EudraCT	2019-001550-26

Document ID: VV-CLIN-081539

Table of Contents

List of Abbreviations

Term	Definition
AC	assessment committee
ADA	anti-drug antibody
AE	adverse event
AESI	adverse event of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
BMI	body mass index
BP	blood pressure
CAP	controlled attenuation parameter
CEC	clinical endpoint committee
CI	confidence interval
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CLDQ-NAFLD	Chronic Liver Disease Questionnaire for Nonalcoholic Fatty Liver Disease
COVID-19	coronavirus disease 2019
CRF	case report form
CSR	clinical study report
DILI	drug-induced liver injury
EAS	efficacy analysis set
ECG	electrocardiogram
eCRF	electronic case report form
EXAS	exploratory analysis set
FAS	full analysis set
GGT	gamma-glutamyl transferase

GI	gastrointestinal
HLT	High Level Term
ICE	intercurrent event
ISR	injection site reaction
LLT	Low Level Term
LSM	liver stiffness measurement
LY	LY3298176
MACE	major adverse cardiovascular event
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intent to treat
MMRM	mixed model for repeated measures
MRE	magnetic resonance elastography
MRI	magnetic resonance imaging
MRI-PDFF	magnetic resonance imaging proton density fat fraction
NAFLD	nonalcoholic fatty liver disease
NAS	NAFLD Activity Score
NASH	nonalcoholic steatohepatitis
NONMEM	Nonlinear Mixed Effects Modeling
PD	pharmacodynamic(s)
PGIS	Patient Global Impression of Severity
PK	pharmacokinetic(s)
PRO	patient-reported outcome
PT	Preferred Term
QW	once weekly
REML	restricted maximum likelihood
SAE	serious adverse event
SAP	statistical analysis plan

SAS	statistical analysis software
SC	subcutaneous
SD	standard deviation
SMQ	standardized MedDRA query
SOC	System Organ Class
SS	safety analysis set
T2DM	type 2 diabetes mellitus
TBL	total bilirubin
TE	transient elastography
TE ADA	treatment-emergent anti-drug antibody
TEAE	treatment-emergent adverse event
TZD	thiazolidinedione
UACR	urine albumin-to-creatinine ration
ULN	upper limit of normal

Version History

SAP Version 1 was approved prior to the first unblinding of the study at the time of the interim analysis.

Table 1 SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	25 April 2022	Not Applicable	Original version
	See date on Page 1		Correct typos Replace CMH with logistic regression Add MRE objectives and analysis Update the analysis set and estimand definition Clarify methods to deal with missing data

1. Study Objectives

1.1. Primary Objective

The primary objective of the study is to demonstrate that tirzepatide 5 mg, 10 mg, or 15 mg administered SC QW is superior to placebo for NASH resolution with no worsening of fibrosis with respect to proportion of participants classified with absence of NASH with no worsening of fibrosis on liver histology at 52 weeks.

1.2. Secondary Objectives

The secondary objectives are to demonstrate that tirzepatide 5 mg, 10 mg, or 15 mg administered SC QW is superior to placebo at Week 52 for:

- Regression of fibrosis with no worsening of NASH with respect to proportion of participants with ≥1-point decrease in fibrosis stage with no worsening of NASH on liver histology from baseline.
- Prevention of fibrosis progression measured with respect to proportion of participants with ≥1-point increase in fibrosis stage on liver histology from baseline.
- Decreasing NAFLD Activity Score (NAS) by ≥2 points with respect to proportion of participants that achieve a ≥2-point decrease in NAS on liver histology, with ≥1-point reduction in at least 2 NAS components (steatosis, hepatocellular ballooning, lobular inflammation) from baseline. The NAS is based on the definition proposed by the NASH Clinical Research Network. The score is defined as the unweighted sum of the scores for steatosis (0-3), lobular inflammation (0-3), and ballooning (0-2); thus ranging from 0 to 8 (Kleiner et al. 2005).
- Decreasing liver fat content with respect to mean absolute change in liver fat content by MRI-PDFF.
- Decreasing body weight with respect to mean change in body weight from baseline.

1.3. Tertiary/Exploratory Objectives

The tertiary exploratory objectives are to investigate the effect of tirzepatide 5 mg, 10 mg, or 15 mg SC QW versus placebo for:

- Additional histological measures with respect to mean change in the liver histology score from baseline for:
    - Expanded ballooning score
    - Expanded portal inflammation score
    - Expanded fibrosis Stage 1 scores
    - Mallory-Denk body score
    - Glycogenosis score

- Liver inflammation and fibrosis with respect to mean change from baseline in extracellular hepatic water content measured by iron-corrected T1 imaging (cT1, in ms) by MRI.
- Liver stiffness with respect to mean change from baseline in liver stiffness (in kPa) measured by TE.
- Liver enzymes and serum biomarkers of NASH and fibrosis with respect to mean changes from baseline in ALT, AST, GGT, K-18, Pro-C3, ELF, FIB-4, Adiponectin, Leptin, and Ferritin.
- Insulin sensitivity with respect to mean changes from baseline in fasting insulin and HOMA-IR.
- Decreasing waist circumference with respect to mean change from baseline in waist circumference (centimeters).
- PROs with respect to changes from baseline in:
    - PROMIS Fatigue short form 8a v1.0 score
    - PROMIS Pain Interference short form 4a v1.0 score
    - CLDQ-NAFLD total and domain scores
    - PGIS
- NASH resolution from baseline based on overall assessment of histology by pathologists.
- Decreasing NAS with respect to mean change from baseline in NAS on liver histology.
- Decreasing steatosis, hepatocellular ballooning, or lobular inflammation with respect to mean change from baseline in the liver histology score for:
    - Steatosis
    - Hepatocellular ballooning
    - Lobular inflammation
- Decreasing steatosis, hepatocellular ballooning, or lobular inflammation by ≥1 point with respect to proportion of participants that achieve a ≥1-point improvement in the liver histology score from baseline for:
    - Steatosis
    - Hepatocellular ballooning
    - Lobular inflammation
- Decreasing fibrosis stage with respect to mean change from baseline in fibrosis stage on liver histology.
- Regression of fibrosis with no worsening of NASH with respect to proportion of participants with ≥2-point decrease from baseline in fibrosis stage, and no worsening of NASH on liver histology.

Approved on 25 Jan 2024 GMT

- Prevention of progression to cirrhosis with respect to proportion of participants classified with cirrhosis (F4) on liver histology.
- Resolution of fibrosis with respect to proportion of participants classified with absence of fibrosis on liver histology.
- Relative decreases in liver fat content with respect to mean percentage change from baseline in liver fat content by MRI-PDFF.
- Decreasing in liver fat content with respect to mean absolute change from baseline in liver fat content by MRI-PDFF.
- Decreasing liver fat content by ≥30% from baseline with respect to proportion of participants that achieve a ≥30% relative decrease in liver fat content by MRI-PDFF compared to baseline.
- Relative decreases in body weight with respect to mean percentage change from baseline in body weight.
- Decreasing body weight by 5%, 10%, 15%, and 20% with respect to proportion of participants that achieve ≥5%, ≥10%, ≥15%, or ≥20% decrease in body weight compared to baseline.

1.4. Magnetic Resonance Elastography (MRE) Addendum Objectives

1.4.1. MRE Primary Objective

- The primary objective is to demonstrate that tirzepatide administered SC QW (pooled data from 5 mg, 10 mg, and 15 mg dose arms) is superior to placebo for the changes from baseline in liver stiffness as measured by 2-dimensional MRE at 52 weeks in participants with NASH.

1.4.2. MRE Secondary Objective

- The secondary objective is to compare tirzepatide, using data pooled from 5 mg, 10 mg, and 15 mg dosing arms, versus placebo on the relative change from baseline in liver stiffness as measured by 2-dimensional MRE at 52 weeks in participants with NASH.

1.4.3. MRE Exploratory Objectives

The exploratory objectives are:

- To compare each dose of tirzepatide (5 mg, 10 mg, and 15 mg) versus placebo on the change from baseline in liver stiffness measured by MRE at Week 26 and Week 52.
- To compare each dose of tirzepatide (5 mg, 10 mg, and 15 mg) versus placebo on the relative change from baseline in liver stiffness measured by MRE at Week 26 and Week 52.

Additional exploratory objectives are to evaluate for tirzepatide, pooled and/or for each individual dose, versus placebo in:

- The proportion of participants achieving ≥15% reduction in liver stiffness from baseline.
- The proportion of participants with ≥15% increase in liver stiffness from baseline.
- Among participants that achieve ≥15% reduction in liver stiffness from baseline, the proportion of participants that achieve ≥1 stage regression of fibrosis on liver histology.
- Among participants with ≥15% increase in liver stiffness from baseline, the proportion of participants with ≥1 stage worsening of fibrosis on liver histology.
- Among participants with ≥15% decrease in liver stiffness from baseline, the proportion of participants achieving NASH resolution on liver histology.

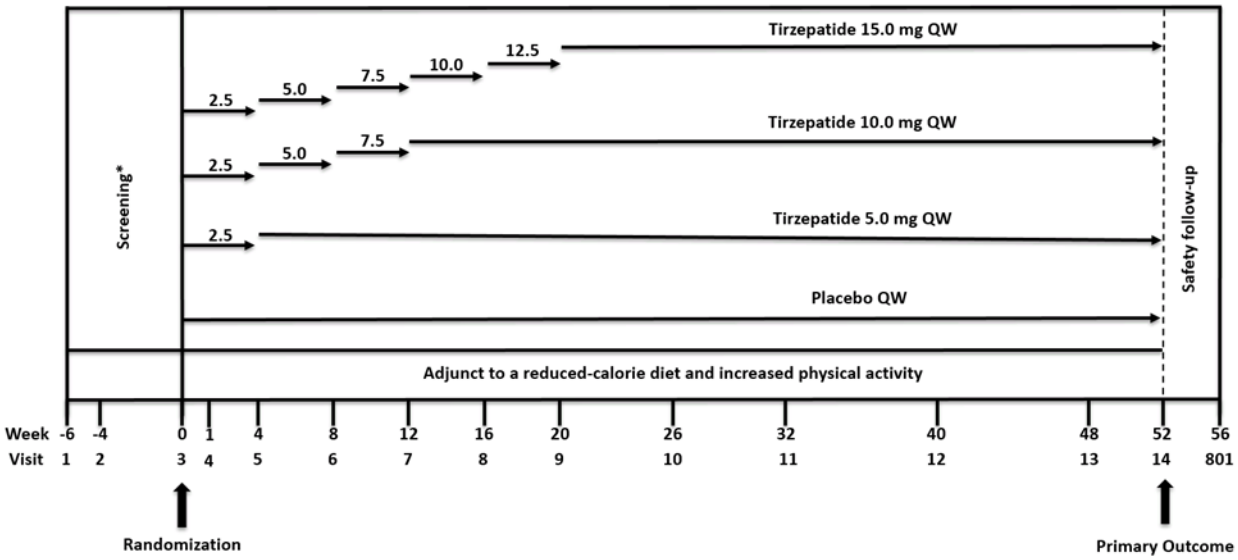
The MRE primary and secondary objectives will also be assessed at Week 26.

2. Study Design

2.1. Summary of Study Design

Study I8F-MC-GPHR (GPHR) is a Phase 2, multicenter, randomized, double-blinded, placebo-controlled study that will investigate the effects of treatment with tirzepatide 5 mg, 10 mg, and 15 mg SC QW compared with placebo on NASH resolution with no worsening of fibrosis in patients with biopsy-proven NASH and fibrosis, with or without T2DM.

Figure GPHR.2.1 illustrates the study design.



Abbreviation: QW = once weekly.

* Screening procedures may take longer or shorter than 6 weeks and will not be considered a protocol deviation.

Figure GPHR.2.1. Illustration of study design for Clinical Protocol I8F-MC-GPHR.

Study GPHR will consist of 3 periods:

- Screening (Visits 1 and 2)
- Treatment (Visits 3 to 14)
  - 4- to 20-week dose escalation (4-week for 5-mg group, 12-week for 10-mg group, 20-week for 15-mg group)
  - 32- to 48-week maintenance (48-week for 5-mg group, 40-week for 10-mg group, 32-week for 15-mg group)
- Safety follow-up (4 weeks post-treatment, through Visit 801)

The maximum total duration of the combined treatment periods is 52 weeks.

Enrichment Strategy

Participants with biopsy-proven NASH will be enrolled into this study. The study population will be enriched for participants who have T2DM, such that 40% to 65% of the total study population will consist of participants with NASH who have also been diagnosed with T2DM. The other 35% to 60% of the study population will consist of participants with NASH who do not have T2DM.

Participant Visit Scheme

Participants with a qualifying liver biopsy performed within 6 months of screening who meet the other required inclusion and exclusion criteria can participate in the study after evaluation of the histology by 2 central pathologists with experience in NASH to confirm presence of NASH and the study-specific histological inclusion criteria. Visit 2 will be omitted for these participants.

Participants without a qualifying liver biopsy will undergo assessments of their CAP and LSM at Visit 1 using TE to determine if they have a high probability of fatty liver disease with fibrosis. Participants meeting prespecified criteria for a high probability of fatty liver disease with fibrosis by TE will undergo a liver biopsy at Visit 2 to assess study eligibility, provided they meet all other required inclusion and exclusion criteria. Liver biopsies will be evaluated by 2 central pathologists with experience in NASH prior to Visit 3 to confirm presence of NASH and study-specific histological inclusion criteria.

2.2. Determination of Sample Size

Sample size selection is guided by the objective of establishing superiority of each tirzepatide dose to placebo relative to the proportion of participants achieving NASH resolution without worsening of fibrosis at 52 weeks from randomization. The evaluation of superiority to placebo will be conducted for each of the 3 tirzepatide doses at 2-sided significance level of 0.05 using a test of 2 proportions. The power is assessed based on the following assumptions:

- the tirzepatide group mean response rate is assumed to be 42.5%
- the placebo group mean response rate is assumed to be 12.5%
- the dropout rate is assumed to be a 20%, and
- 1:1:1:1 randomization.

On the basis of these assumptions, approximately 196 participants will be required to provide at least 80% power to establish superiority of tirzepatide 5-mg, tirzepatide 10-mg, or tirzepatide 15-mg doses to placebo. This results in approximately 40 completers per arm. No adjustment for multiplicity will be performed.

2.3. Method of Assignment to Treatment

Approximately 196 patients who meet all criteria for enrollment will be randomized to one of the study treatment arms at Visit 3. Assignment to treatment arms will be determined by a computer-generated random sequence using an interactive web response system. Patients will be randomized in a 1:1:1:1 ratio to receive 5-mg tirzepatide, 10-mg tirzepatide, 15-mg tirzepatide, or placebo. Patients will be stratified at randomization based on T2DM status (Yes or No) and the region (Japan, US including Mexico, and Europe including Israel).

3. A Priori Statistical Methods

3.1. Population for Analyses

For purposes of analysis, Table GPHR.3.1 defines the following analysis populations/data sets:

Table GPHR.3.1. Analyses Populations/Data Sets

Population/Data Set	Description
Entered population	All participants who sign informed consent
Randomized population	All participants who are randomly assigned to a treatment arm
Modified intent-to-treat (mITT) population	All randomly assigned participants who are exposed to at least 1 dose of study drug. In the event of a treatment error, participants will be analyzed according to the treatment they were randomized.
Efficacy analysis set (EAS)	Data obtained during treatment period from all randomized participants, excluding data after permanent study drug discontinuation. Participants will be analyzed according to the treatment they were randomized.
Exploratory analysis set (EXAS)	Data obtained during treatment period from mITT population participants who stay on study treatment for at least 36 weeks, excluding data after permanent study drug discontinuation.
Full analysis set (FAS)	Data obtained during treatment period from the mITT population, regardless of adherence to study drug.
Safety analysis set (SS)	Data obtained during treatment period plus the safety follow-up period from the mITT population, regardless of adherence to study drug.

3.2. General Considerations

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly) or its designee. All statistical analyses will be conducted with SAS Version 9.4 or higher unless otherwise stated. Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (e.g., few events to justify conducting an analysis). Listings of events will be provided in such situations. Additional analyses of the data may be conducted as deemed appropriate without further changes made to the protocol or SAP, even after the primary or final database locks.

Additionally, to avoid potential selection biases, unless stated otherwise, statistical summaries and analyses will be conducted based on randomized study treatment group regardless of the actual treatment received by the patient. Unless otherwise noted, all 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. In statistical summaries and analyses, patients will be analyzed as randomized.

For histological, imaging, and TE measurements the last available data during Visit 1 to Visit 3 prior to or within 1 week after the first study drug administration will serve as baseline. For patients who have an eligible historical biopsy within 6 months of Visit 1, the historical biopsy will serve as baseline for histological measures. Unless specified otherwise, the last measurement during Visit 1 to Visit 3 (including unscheduled visits) collected prior to or on the first dose day will serve as baseline. For immunogenicity, data collected up to the first dose time will serve as baseline. For laboratory and ECG, baseline needs to be prior to or within 1 hour after the first dose time. For PROs, data collected at Visit 3, regardless of the timing relative to the first dose, will serve as baseline.

There will be 2 estimands of interest in evaluating the primary and secondary efficacy objectives. The first estimand, the "efficacy" estimand, represents efficacy prior to discontinuation of study drug.

Primary Estimand (efficacy estimand)

The primary clinical question of interest is: What is the treatment difference in NASH resolution without worsening of fibrosis after 52 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 Section 5 of Protocol GPHR.
- Endpoint: NASH resolution without worsening of fibrosis at 52 weeks.
- Treatment condition: the randomized treatment.
- 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.
- Population-level summary: proportion of patients meet NASH resolution without worsening of fibrosis at Week 52.

Treatment effect of interest is thus defined as the difference in proportion of NASH resolution without worsening of fibrosis at Week 52 between QW LY3298176 (tirzepatide) and placebo. Rationale for "efficacy estimand": This Phase 2 study aims to study the efficacy of LY3298176 (tirzepatide) under the ideal condition that all participants adhere to the randomized treatment. Data collected after the ICE will be set to missing.

Estimand(s) for Secondary Objectives

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

Treatment-regimen estimand

The second estimand, the "treatment-regimen" estimand, represents the efficacy irrespective of adherence to investigational product and availability of the postbaseline biopsy. There are no ICE.

Unless specified otherwise, safety analyses will be conducted using the SS.

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

Summary statistics for categorical measures (including categorized continuous measures) will include sample size, frequency, and percentages. Summary statistics for continuous measures will include sample size, mean, SD, median, minimum, and maximum. The summary statistics

will be presented by nominal visit. The analysis model to make comparisons among treatment groups relative to continuous measurements assessed over time will be either ANCOVA or an MMRM.

Statistical treatment comparisons will be performed between tirzepatide doses and placebo. The comparison among tirzepatide groups will be performed if deemed necessary.

Statistical summaries and results of statistical analyses will be displayed in the following treatment order: 5-mg tirzepatide, 10-mg tirzepatide, 15-mg tirzepatide, placebo.

Not all analyses described in this SAP will necessarily be included in the CSR. Any analysis described in this SAP and not provided in the CSR would be available upon request.

3.3. Adjustments for Covariates

The study is stratified based on T2DM status (Yes or No) at randomization and the region (Japan, US including Mexico, and Europe including Israel).

For efficacy-related analyses of histological endpoints, region and T2DM status at randomization will be used as covariates. Unless otherwise specified, for efficacy analyses of continuous measurements including non-histological and laboratory biomarkers, corresponding baseline value will be used as an additional covariate to T2DM status and the region.

3.4. Handling of Dropouts or Missing Data

For the analyses of primary and key secondary efficacy endpoints related to histology, data for participants with missing postbaseline liver biopsy will be excluded and not be imputed. A separate analysis for these endpoints may be conducted using FAS, where missing postbaseline biopsies will be imputed based on the method described in Section 3.11.1.3.

Unless specified otherwise, imputation of missing data for exploratory purposes will be limited to primary and key secondary efficacy endpoint analyses.

3.5. Multicenter Studies

To investigate potential regional influence on efficacy, region (Japan, US including Mexico, and Europe including Israel) will be used as a stratification factor in primary and secondary efficacy analysis.

3.6. Multiple Comparisons/Multiplicity

As this is a phase 2 exploratory study, no multiplicity adjustment will be made for the analyses.

3.7. Patient Disposition

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

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

3.8. Patient Characteristics

A listing of patient demographics will be provided for all randomized patients. All demographic and baseline clinical characteristics will be summarized by study treatment for all randomized patients. Baseline demographic and clinical characteristics of special interest include but not limited to:

- age (years)
- sex (female, male)
- race
- ethnicity
- weight (kg)
- BMI (kg/m^2)
- waist circumference (cm)
- T2DM status (yes, no)
- country,
- liver fat content (%)
- NAS score
- fibrosis stage
- steatosis
- lobular inflammation
- hepatocellular ballooning
- ALT
- AST
- HbA1c
- CAP (TE), and
- LSM (TE).

3.9. Treatment Exposure and Compliance

A listing of patients randomized but not receiving study treatment will be provided, if applicable. The listing will include patient identification, randomized treatment arm, and the reason for not receiving study treatment, if the data is available.

A listing of randomized patients who inadvertently received incorrect study treatment anytime during the study will be provided, if applicable. The listing will include patient identification, randomized treatment arm, and information related to the treatment incorrectly received: incorrectly received study treatment, start and stop dates during which the incorrect treatment was received.

Summary of duration of follow-up (defined as time in days from date of randomization to date of the safety follow-up, date of early study discontinuation or date of death) and 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.

3.9.1. Compliance to Study Treatment

The number of patients prematurely discontinuing study treatment prior to the 52-week visit will be provided by study treatment arms. Reasons for prematurely discontinuing study treatment prior to the 52-week visit will be provided by study treatment. Time-to-event analysis of premature study treatment discontinuation will be conducted.

If data warrants, the counts and percentages of participants who follow the planned escalation scheme, have dose interruption, or have dose de-escalation will be summarized for each treatment group. In addition, the proportion of participants with missing dosing information, receiving no LY dose, receiving 2.5 mg, 5 mg, 7.5 mg, 10 mg, 12.5 mg, or 15 mg may be presented by randomized treatment and time intervals from first dose.

Overall treatment compliance will be defined as taking at least 75% of the scheduled tirzepatide doses. Compliance will be calculated by taking the number of doses administered (regardless of the actual dose in mg administered) divided by the total number of doses expected to be administered, and then multiplied by 100. Overall treatment compliance will be summarized descriptively by treatment using the mITT population.

3.10. Concomitant Therapy

Prespecified concomitant medications of interest will be summarized by treatment at randomization. Additionally, medications of interest initiated after randomization will be summarized. Concomitant therapies will be mapped using the World Health Organization Drug Dictionary in the clinical trial database and will be further classified using Anatomic-Therapeutic-Chemical codes for reporting purposes.

Concomitant medication summaries of interest include:

- baseline antihypertensive therapy, by type
- baseline lipid lowering therapy, by type
- baseline glucose lowering medications (including TZDs)
- baseline vitamin E
- changes to baseline medication in post randomization (in term of type/class and dose):
    - antihypertensive therapy
    - lipid lowering therapy
    - glucose lowering medications (including TZDs)
    - vitamin E
- utilization after randomization of:
    - antihyperglycemic rescue therapy for hyperglycemia for participants who have T2DM at baseline, and antihyperglycemic medication for the treatment of diabetes for participants who develop T2DM during the study (antihyperglycemic medication for the treatment of prediabetes is not allowed pre protocol)
    - antidiarrheal medication, and
    - antiemetic medication.

3.11. Efficacy Analyses

The assessment of the primary, secondary and exploratory endpoints will be guided by the "efficacy" estimand. Analyses relative to the "efficacy" estimand will be conducted using the EAS without imputation of missing data. For exploratory purposes, and in the event of patients dropping out before Week 36, primary and key secondary efficacy analyses may be repeated using the EXAS guided by the "efficacy" estimand.

For informational purposes, the analysis of primary and key secondary endpoint may be repeated to be guided by the "treatment-regimen" estimand, which represents the efficacy irrespective of adherence to investigational product. Analyses relative to the "treatment-regimen" estimand will be conducted using the FAS. In this case, missing postbaseline measures will be imputed according to Section 3.11.1.3.

3.11.1. Primary Efficacy Analysis

The primary efficacy measure will be proportion of participants classified with absence of NASH with no worsening of fibrosis on liver histology at 52 weeks. The resolution of NASH is obtained when **all 4** conditions in the following are satisfied:

- absence of fatty liver disease (NAS 0 for steatosis) or simple steatosis without steatohepatitis
- the absence of hepatocellular ballooning (NAS 0 for ballooning)
- with or without mild lobular inflammation (NAS 0 or 1 for inflammation), and
- any value for steatosis.

No worsening of fibrosis will be defined as no increase in fibrosis stage from baseline to Week 52. The proportion of participants with NASH resolution without worsening of fibrosis will be summarized by treatment at 52 weeks.

3.11.1.1. The Analysis Relative to the Efficacy Estimand

The analysis related to efficacy estimand will be conducted utilizing the NASH resolution with no worsening of fibrosis (Yes/No) at 52 weeks in the EAS. Logistic regression will be used to evaluate performance of each tirzepatide arm compared to placebo. The test will adjust for stratification factors by including baseline T2DM status (Yes/No) and the region (Japan, US including Mexico, and Europe including Israel).

With the aid of the logistic regression, p-values of the test as well as 2-sided 95% CI of the odds ratio for the proportion of patients achieving NASH resolution with no worsening of fibrosis at 52 weeks between each dose of tirzepatide 5 mg, 10 mg, and 15 mg and placebo will be derived.

3.11.1.2. The Analysis Relative to the Treatment-Regimen Estimand

The analysis related to treatment-regimen estimand may be conducted utilizing the NASH resolution (Yes/No) with no worsening of fibrosis at 52 weeks in the FAS. The same logistic regression described in Section 3.11.1.3 will be used for the analysis. Multiple imputation of missing NASH resolution status will be conducted before performing the logistic regression (see Section 3.11.1.3).

3.11.1.3. Methods for Multiple Imputations

For efficacy analyses relative to "treatment-regimen" estimand, the ICEs and the resulting missing values will be handled as follows:

There are no ICE for "treatment-regimen" estimand. If the endpoint biopsy is not collected, data from participants who discontinue, but with non-missing endpoint biopsy from the same treatment arm will be used to impute the missing value (retrieved dropout imputation). In cases where there are not enough retrieved dropouts to provide a reliable imputation model, the imputation of missing data will be done with the jump-to-reference (placebo) imputation approach.

For NASH resolution with no worsening of fibrosis which is a binary endpoint and multiple conditions (see Section 3.11.1) need to be satisfied in addition to no increase in fibrosis score, the imputation will be done for each of the components separately. That is, ballooning score, steatosis score, inflammation score, and fibrosis score will be imputed separately. At the end, the imputed values will determine if the participant satisfies all criteria to have "Yes" for NASH resolution with no worsening of fibrosis or "No" otherwise based on the criteria listed in Section 3.11.1.

The statistical inference over multiple imputation of missing data is guided by methodology proposed by Rubin (1987).

3.11.1.4. Additional Analyses of the Primary Outcome

In the event of patients dropping out before Week 36, primary analyses may be repeated using the EXAS for exploratory purposes. The model and details of analysis is similar to what is described in Section 3.11.1.1.

3.11.2. Secondary Efficacy Analyses

3.11.2.1. Analyses of Continuous Outcomes

Analyses for change from baseline in body weight as well as mean absolute change in liver fat content by MRI-PDFF will be conducted using data in the EAS from baseline through the 52-week visit with the aid of an MMRM. REML will be used to obtain model parameter estimates and the Kenward-Roger option will be used to estimate denominator degrees of freedom. The response variable of the MMRM model will be the primary measure and model terms of interest will include treatment, visit, treatment-by-visit interaction, baseline T2DM status (Yes/No), and region (Japan, US including Mexico, and Europe including Israel) as fixed effects, and baseline values of the dependent variable as a covariate. An unstructured covariance structure will be used to model the within-patient errors. If this model fails to converge, the following covariance structures will be tested in the following order:

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

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

If the outcome has only the baseline and end-of-treatment measurements, the analysis will be conducted utilizing ANCOVA. The response variable will be the continuous outcome and model terms will include treatment, baseline T2DM status (Yes/No), and region (Japan, US including Mexico, and Europe including Israel) as fixed effects, and baseline values of the outcome as a covariate.

3.11.2.2. Analyses of Binary Outcomes

Analyses for proportion of participants with ≥1-point decrease in fibrosis stage with no worsening of NASH on liver histology Week 52 from baseline will be conducted using the same statistical models as those used for evaluating the primary objective in Section 3.11.1.1. No worsening of NASH is defined as no increase in the NAS score.

Analyses for proportion of participants with ≥1-point increase in fibrosis stage as well as proportion of participants that achieve a ≥2-point decrease in NAS on liver histology, with ≥1-point reduction in at least 2 NAS components (steatosis, hepatocellular ballooning, lobular inflammation) at Week 52 from baseline will be conducted using the same statistical models as those used for evaluating the primary objective in Section 3.11.1.1.

3.11.2.3. The Analysis Relative to the Treatment-Regimen Estimand

The analysis related to treatment-regimen estimand for all secondary outcomes may be conducted using the FAS. The same model described in Section 3.11.2.2 will be used for the analysis of binary outcomes. Missing values of the fibrosis score and components of NAS will be imputed according to Section 3.11.1.3.

For the mean body weight and mean absolute change in liver fat content by MRI-PDFF at Week 52, the analyses will be conducted utilizing ANCOVA described in Section 3.11.2.1. Imputation of missing data will be done with the jump-to-reference (placebo) imputation approach.

3.11.2.4. Additional Analyses of the Secondary Outcomes

Similar to primary outcomes, in the event of patients dropping out before Week 36, analyses may be repeated using the EXAS for exploratory purposes. The model and details of analysis is similar to what is described in Section 3.11.1.1.

3.11.3. Tertiary/Exploratory Efficacy Analyses

Unless otherwise specified, other secondary and exploratory efficacy analyses will be conducted using the EAS. Missing data will be handled through MMRM without utilizing multiple imputation technique (if applicable). Some biomarkers may be log transformed, if necessary.

3.11.3.1. Analyses of histological scores

Parts of the exploratory efficacy analyses include analyzing scores obtained from biopsy reads (e.g., components of NAS, fibrosis score, etc.). ANCOVA is utilized to perform the analysis in

such. Treatment, baseline T2DM status (Yes/No), and region (Japan, US including Mexico, and Europe including Israel) are the fixed effects in this model.

Table GPHR.3.2 describes appropriate analysis for tertiary/exploratory outcome with the objective of investigating the effect of tirzepatide 5 mg, 10 mg, or 15 mg SC QW versus placebo at different time points specified for each efficacy measure separately.

Table GPHR.3.2. Tertiary/Exploratory Efficacy

Efficacy Measure	Analysis Conducted	Additional Information
Mean change in the liver histology score for • expanded ballooning score • expanded portal inflammation score • expanded fibrosis Stage 1 scores • Mallory-Denk body score, and • glycogenosis score	ANCOVA in Section 3.11.2.1	Each of these items come from each pathologist's eCRF form and are not in the consensus form. As a result, the mean values from each pathologist will be used as the response variable.
Mean change in extracellular hepatic water content measured by iron-corrected T1 imaging (cT1, in ms) by MRI	MMRM model in Section 3.11.2.1	
Mean change in liver stiffness (in kPa) measured by TE	MMRM model in Section 3.11.2.1	
Mean change in ALT, AST, GGT, K-18, Pro-C3, ELF, FIB-4, adiponectin, leptin and ferritin	MMRM model in Section 3.11.2.1	ALT, AST, and GGT need to be log-transformed. Log-transformation may be applied to other biomarkers as appropriate.
Mean changes in fasting insulin and HOMA-IR	MMRM model in Section 3.11.2.1	Log-transformation may be applied as appropriate.
Mean change in waist circumference (centimeters)	MMRM model in Section 3.11.2.1	
Change in: • PROMIS Fatigue Short Form 8a version 1.0 score • PROMIS Pain Interference Short Form 4a version 1.0 score • Chronic Liver Disease Questionnaire for Nonalcoholic Fatty Liver Disease (CLDQ-NAFLD) total and domain scores • Patient Global Impression of Severity (PGIS)	ANCOVA model in Section 3.11.2.1	Psychometric analyses on PROs will be conducted by value, evidence, and outcomes (VEO) at Lilly
NASH resolution based on overall assessment of histology by pathologists	Logistic regression in Section 3.11.1.1	This directly comes from the pathologists consensus form in the CRF regarding determination if steatohepatitis is present.
Mean changes in NAS on liver histology	ANCOVA in Section 3.11.3.1	NAS is the sum of Steatosis, Lobular Inflammation, and Hepatocellular Ballooning scores directly from the pathologists consensus form in the CRF.

Efficacy Measure	Analysis Conducted	Additional Information
Mean change in the liver histology score for • steatosis • hepatocellular ballooning, and • lobular inflammation	ANCOVA in Section 3.11.3.1	Each of these items come directly from the pathologist's consensus form in the CRF.
Proportion of participants that achieve ≥1-point improvement in the liver histology score for • steatosis • hepatocellular ballooning, and • lobular inflammation	Logistic regression in Section 3.11.1.1	Each of these items come directly from the pathologist's consensus form in the CRF.
Mean change in fibrosis stage on liver histology	ANCOVA in Section 3.11.3.1	This comes directly from the pathologist's consensus form in the CRF.
Proportion of participants with ≥2-point decrease in fibrosis stage and no worsening of NASH on liver histology	Logistic regression in Section 3.11.1.1	No worsening of NASH is defined as no increase in the NAS score.
Proportion of participants classified with cirrhosis (F4) on liver histology	Logistic regression in Section 3.11.1.1	
Proportion of participants classified with absence of fibrosis on liver histology	Logistic regression in Section 3.11.1.1	
Mean percentage change in liver fat content by MRI-PDFF	MMRM model in Section 3.11.2.1	The analysis is done on relative changes in liver fat content at each measurement.
Proportion of participants that achieve a ≥30% relative decrease in liver fat content by MRI-PDFF	Logistic regression in Section 3.11.1.1	The analysis should be done for Week 26 and Week 52 separately.
Mean percentage change in body weight	MMRM model in Section 3.11.2.1	
Proportion of participants that achieve ≥5%, ≥10%, ≥15%, or ≥20% decrease in body weight	Longitudinal logistic regression model with the response variable of the percentage of participants achieving at least those percentages of body weight reduction at each scheduled postbaseline visit	The independent variables of the analysis model are treatment group, visit, treatment-by-visit interaction, and stratification factors (T2DM status and Region) as fixed effects and baseline body weight as a covariate.

Abbreviations: ALT = alanine aminotransferase; ANCOVA = analysis of covariance; AST = aspartate aminotransferase; CRF = case report form; eCRF = electronic CRF; ELF = Enhanced Liver Fibrosis panel; FIB-4 = Fibrosis-4 score; GGT = gamma-glutamyltransferase; HOMA-IR = Homeostatic Model Assessments for Insulin Resistance; K-18 = Keratin-18 M30 fragment; Lilly = Eli Lilly and Company; MMRM = mixed model for repeated measures; MRI = magnetic resonance imaging; NAS = NAFLD Activity Score; NASH = nonalcoholic steatohepatitis; PDFF = proton density fat fraction; PRO = patient-reported outcome; Pro-C3 = a fragment of the NH_2-terminal propeptide of type III procollagen; PROMIS = Patient Reported Outcome Measurement Information System; TE = transient elastography; T2DM = type 2 diabetes mellitus.

3.12. Health Outcomes

The PRO 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 will be performed post-hoc by Value, Evidence, and Outcomes at Lilly and documented in a separate analysis plan.

3.12.1. Patient Global Impression of Severity (PGIS)

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

3.12.2. Chronic Liver Disease Questionnaire for Nonalcoholic Fatty Liver Disease (CLDQ-NAFLD)

The following parameters will be included from the CLDQ-NAFLD:

- Abdominal domain (3 items; Items 1, 5, 17)
- Activity domain (5 items; Items 7, 9, 14, 30, 31)
- Emotion domain (9 items; Items 10, 12, 15, 16, 19, 20, 24, 26, 34)
- Fatigue domain (6 items; Items 2, 4, 8, 11, 13, 35)
- Systemic domain (6 items; Items 3, 6, 21, 23, 27, 36)
- Worry domain (7 items; Items 18, 22, 25, 28, 29, 32, 33)
- Total score

Individual items are rated on a 7-point scale ranging from “1- All of the time” to “7- None of the time.” Scores are calculated separately for each domain by taking the average of the domain’s items. The total score is calculated by taking the average of the domain scores. Higher scores reflect better health. In the case of missing values, domains scores can be calculated when at least 50% of the items are completed, and the total score should include at least 4 domains.

3.12.3. PROMIS Short Form Fatigue 8a v1.0

The counts and percentages of participants endorsing each response category for each item of the PROMIS Short Form Fatigue 8a v1.0 at each time point will be summarized for each nominal visit and by treatment. Individual item scores are totaled to obtain a raw score, with higher scores indicating more interference. Raw scores can be converted to a T-score.

3.12.4. PROMIS Short Form Pain Interference 4a v1.0

The counts and percentages of participants endorsing each response category for each item of the PROMIS Short Form Pain Interference 4a v1.0 at each time point will be summarized for each nominal visit and by treatment. Individual item scores are totaled to obtain a raw score, with higher scores indicating more interference. Raw scores can be converted to a T-score.

3.13. Pharmacokinetic/Pharmacodynamic Methods

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

Approved on 25 Jan 2024 GMT

A summary of tirzepatide concentration-time data will be reported in the CSR. Exposure-response analysis between tirzepatide concentration and safety, pharmacology, and efficacy may be performed using population PK and population PK/PD nonlinear mixed-effects modeling techniques implemented on NONMEM software.

If ADA titers are detected in the immunogenicity samples, the impact of titers on tirzepatide PK and/or PD may be evaluated.

3.14. Safety Analyses

Unless specified otherwise, safety assessments will be based on the SS (Table GPHR.3.1). All events that occur between the date of first dose of study drug to the date of patient's safety follow-up visit or patient's end of study participation will be included, regardless of the adherence to study drug or availability of the end of treatment liver biopsy. For assessing the benefit and risk profile through 52 weeks, selected safety analyses will be conducted by utilizing safety data from first dose through the date of the 52-week visit.

The statistical assessment of homogeneity of the distribution of categorical safety responses among treatment arms will be conducted using Fisher's exact test, unless specified otherwise.

The mean change from baseline differences among treatments at all scheduled visits will be assessed via an MMRM using REML, unless specified otherwise. The model will include baseline T2DM status (Yes/No), the region (Japan, US including Mexico, and Europe including Israel), treatment group, visit and treatment-by-visit interaction as fixed effects, and baseline value of the safety parameter as a covariate. To model the covariance structure within patients, the unstructured covariance matrix will be used. If this model fails to converge, the covariance structures specified in Section 3.11.2.1 will be tested in order.

For selected safety parameters, time-to-first-event analysis via the Cox Proportional-Hazards model may be conducted. Patients without the event will be censored at the end of study participation. For patients experiencing the event, the "time-to-first-event" will be the time (in days) from first dose to first occurrence of the event.

Where necessary, the rate of events will be analyzed using a generalized linear mixed-effects model assuming the number of events follow a negative binomial distribution and with treatment as a fixed effect. The logarithm of days during the active treatment period will be adjusted as an offset to account for possible unequal treatment duration of follow-up between patients.

3.14.1. Adverse Events

If applicable, a listing of AEs occurring either before first dose or after the patient's last date of study participation will be provided. The listing will include patient identification including the treatment, site number, event information: AE group ID, event start date, MedDRA SOC, and PT, seriousness, severity, outcome, relationship to study drug, time from first dose of study drug to the event, time from last dose of study drug to the event, and time from end of study participation to the event.

3.14.1.1. Treatment-Emergent Adverse Events

A TEAE is defined as an event that first occurred or worsened in severity after first dose of study drug. 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 (e.g., treatment-emergent flag, start time of study treatment and event) will be used to determine whether the event was pre- versus post-treatment, if available. If the relevant information is not available, then the events will be counted as post-treatment.

The counts and percentages of participants with TEAEs will be summarized by treatment using MedDRA PT nested within SOC. 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, and outcome of the AE 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.

3.14.1.2. Common Adverse Events

The counts and percentages of participants with TEAEs, overall and common (common TEAEs occurred in ≥1.0% of participants before rounding), will be summarized by treatment using MedDRA PT. Events will be ordered by decreasing frequency.

3.14.1.3. Deaths

A listing of all deaths during the study will be provided. The listing will include participant identification including the treatment, site number, date of death, age at the time of enrollment, sex, MedDRA PT, associated AE group identification, time from last dose of study drug to death (if participant had discontinued study drug), investigator-reported cause of death, and CEC-adjudicated cause of death.

3.14.1.4. Other Serious Adverse Events

The counts and percentages of participants who experienced an SAE (including deaths and SAEs temporally associated or preceding deaths) during the postbaseline period will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC. The SOC will be in alphabetical order.

A listing of all SAEs will be provided. The listing will include, but not be limited to, treatment, participant identification including the site number, date of event, age at the time of enrollment, sex, AE group identification, MedDRA SOC and PT, severity, action taken, outcome, relationship to study drug, time from first dose of study drug to the event, and event duration.

3.14.1.5. Discontinuation Due to Adverse Events

The counts and percentages of participants who discontinued from study treatment or study due to an AE during the postbaseline period may be summarized by treatment group using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC. In case of discontinuation from study treatment, a time-to-event analysis will be conducted by treatment on time to study drug discontinuation as well as on time to study drug discontinuation due to an AE.

3.14.1.6. Treatment of Overdose

A listing of patients reporting AEs related to overdosing of tirzepatide will be provided. Per protocol, any dose of study drug greater than the highest possible dose (15 mg) and estimated from where the patient is in the dose-escalation regimen or treatment-maintenance regimen within a 2-day time period will be considered an overdose and should be reported as an AE.

3.14.2. Patient Narratives

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

- death
- SAE
- pregnancy
- permanent discontinuation of study treatment due to AEs, or
- severe adverse events of special interest.

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

3.14.3. Special Safety Topics

3.14.3.1. Hypoglycemia

Clinically significant hypoglycemia (Level 2 or 3 hypoglycemia) will be summarized as the number of patients with at least one hypoglycemia, incidence, the number of total episodes, and the rate per 100 patient years during the study by treatment group if deemed necessary. Based on ADA (2022), Level 2 hypoglycemia is defined as a glucose less than 54 mg/dL (3.0 mmol/L), and Level 3 hypoglycemia is defined as a severe event characterized by altered mental and/or physical status requiring assistance for treatment of hypoglycemia; it is also referred to as severe hypoglycemia. 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.

Hypoglycemic episodes will be recorded on a specific eCRF. Information regarding severity, time of day, and investigator's opinion of relatedness to study drug and procedure will be listed. Hypoglycemic episodes will be recorded as an AE only if it meets serious criteria.

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

Hypoglycemia will be considered as an AESI. A listing of all events of serious or severe hypoglycemia may be provided, if deemed necessary. This listing will provide treatment allocation, clinical characteristics of the hypoglycemic event, and concomitant antihyperglycemic medications.

3.14.3.2. Hepatic Safety

3.14.3.2.1. NASH-Related Clinical Events

NASH-related clinical events will be considered as AESI. The counts and percentages of participants with the event will be summarized by treatment.

A listing of participants with adjudicated NASH-related clinical events will be provided. The listing will include patient identification including the treatment, site number, date of the event, sex, time from first dose of study drug to the event, time from last dose of study drug to the event (if patient had discontinued study drug), and adjudicated description of the event (if available).

3.14.3.2.2. Hepatobiliary Disorders

Hepatobiliary disorders will be considered as AESI. The counts and percentages of participants with treatment-emergent potentially drug-related hepatic disorders will be summarized by treatment using the MedDRA SOC and PTs. Detailed searching criteria can be found in Appendix 1 (Section 6.1).

A listing of participants with treatment emergent hepatobiliary disorders may be provided if deemed necessary.

3.14.3.2.3. Drug-Induced Liver Injury (DILI)

Adjudicated DILI events will be considered as AESI. The counts and percentages of participants with the event will be summarized by treatment. A listing of participants with DILI events may be provided if deemed necessary.

3.14.3.2.4. Liver Enzymes

Analyses for laboratory analyte measurements are described in Section 3.14.4. 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 treatment period and during the entire study including follow up period will be summarized between treatment groups:

- The counts and percentages of participants with an ALT measurement 1 or more times (1x), 3 or more times (3x), 5 or more times (5x), and 10 or more times (10x) the central laboratory ULN will be summarized for all participants with a postbaseline value and for subsets based on various levels of baseline value.
    - participants whose nonmissing maximum baseline value is ≤1x ULN

  - participants whose maximum baseline is >1x ULN, and
  - participants whose baseline values are missing.
- The counts and percentages of participants with an AST measurement 1 or more times (1x), 3 or more times (3x), 5 or more times (5x), and 10 or more times (10x) the central laboratory ULN during the treatment period will be summarized for all patients with a postbaseline value and for subsets based on various levels of baseline, as described above for ALT.
- The counts and percentages of participants with a TBL measurement 1 or more times (1x), 2 or more times (2x), the central laboratory ULN during the treatment period will be summarized for all participants with a postbaseline value, and for subsets based on various levels of baseline values:
  - participants whose nonmissing maximum baseline value is ≤1 x ULN
  - participants whose maximum baseline is >1 x ULN, and
  - participants whose baseline values are missing.
- The counts and percentages of participants with a serum ALP measurement 1 or more times, 2 or more times (2x) the central laboratory ULN during the treatment period will be summarized for all participants with a postbaseline value and for the subsets described above to TBL.

Maximum baseline will be the maximum non-missing observation in the baseline period. The maximum value will be the maximum non-missing value from the postbaseline period. Planned and unplanned measurements will be included. Evaluation of drug induced serious hepatotoxicity plots will be performed to show the relation between maximum postbaseline ALT (ULN) versus maximum postbasline TBL (ULN), and maximum postbaseline AST (ULN) versus maximum postbaseline TBL (ULN).

3.14.3.3. Exocrine Pancreas Safety

3.14.3.3.1. 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 the ULN, and greater than the ULN), and treatment: less than or equal to 1 times the ULN, greater than 1 to less than or equal to 3 times the ULN, greater than 3 to less than or equal to 5 times the ULN, greater than 5 to less than or equal to 10 times the ULN, greater than 10 times the ULN.

An MMRM analysis will be used to analyze each pancreatic enzyme with a log transformed (postbaseline measure/baseline measure) response variable and treatment, nominal visit, treatment-by-nominal visit interaction as fixed effects.

3.14.3.3.2. Pancreatitis Events

Summaries of adjudicated and investigator-reported pancreatic events will be provided by treatment. Detailed searching criteria can be found in Appendix 1 (Section 6.1).

Treatment-emergent, adjudicated pancreatitis will be considered as AESI. Listing of participants with adjudicated pancreatitis may be provided if deemed necessary.

3.14.3.4. Gastrointestinal Safety

3.14.3.4.1. Nausea, Vomiting, and Diarrhea

The time courses of prevalence and incidence (newly occurring episodes) of nausea, vomiting, diarrhea, and combined will be plotted by treatment and maximum severity. The maximum severity and duration of treatment-emergent nausea, vomiting, diarrhea, and combined through the end of the study will be summarized by treatment.

3.14.3.4.2. 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 gastrointestinal SOC will be used to identify GI AEs. The incidence of the resulting TEAEs will be summarized by treatment and PT. PTs with severe/serious cases in the GI SOC will be considered AESIs. The counts and percentages of participants with severe GI events will be summarized by treatment.

3.14.3.5. Thyroid Safety Monitoring

3.14.3.5.1. Calcitonin

Observed calcitonin data will be summarized by treatment and nominal visit. Additionally, the number and proportion of patients with a maximum postbaseline calcitonin value exceeding the following thresholds will be provided by treatment and maximum baseline calcitonin value (less than or equal to 20 ng/L, greater than 20 ng/L): less than or equal to 20 ng/L, greater than 20 ng/L to less than or equal to 35 ng/L, greater than 35 ng/L to less than or equal to 50 ng/L, greater than 50 ng/L to less than or equal to 100 ng/L, greater than 100 ng/L.

3.14.3.5.2. 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 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.

3.14.3.6. Major Adverse Cardiovascular Events (MACE)

MACE reported by investigators are adjudicated by an independent CEC in a blinded fashion.

The MACE of special interest are: deaths due to cardiovascular 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.

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).

3.14.3.7. Treatment-Emergent Supraventricular Arrhythmias and Cardiac Conduction Disorders

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

The AE database will be searched using predefined SMQs or MedDRA HLTs to identify events consistent with supraventricular arrythmias and cardiac conduction disorders. Detailed searching criteria can be found in Appendix 1 (Section 6.1).

The incidences of the resulting TEAE will be summarized by treatment and PT nested within SMQs and HLTs. 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.

3.14.3.8. Hypersensitivity Reactions

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

- **Potential Immediate Hypersensitivity:** Analysis of TEAEs occurring from the start of study drug administration up to 24 hours after the end of study drug administration. For events without the hypersensitivity eCRF, only date (no time) information is collected, and the events occurring on the same date as the study drug injection 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. Detailed searching criteria for hypersensitivity events can be found in Appendix 1 (Section 6.1).

3.14.3.9. Injection Site Reactions

ISRs, incidence and rates, and related information reported via the “Injection Site Reactions” eCRF will be summarized by treatment. Information to be summarized includes the location of the reaction, timing of the reaction relative to study drug administration, and characteristics of the ISR: erythema, induration, pain, pruritis, and edema.

Additionally, potential ISRs will be reported by PT within MedDRA HLTs of ISRs, administration site reactions, and infusion-related reactions. The counts and percentages of participants with treatment-emergent ISR will be summarized by treatment using the MedDRA

PTs. Detailed searching criteria for ISR events can be found in Appendix 1 (Section 6.1). The PTs will be used for summary within each HLT category. All ISRs will be considered AESIs.

The counts and percentage of participants with ISRs will be summarized by treatment. A listing of participants with treatment-emergent severe/serious ISRs may be provided, if deemed necessary.

3.14.3.10. Diabetic Retinopathy Complications

Any TEAE suspected of worsening retinopathy triggers a follow-up dilated fundoscopic exam. A summary of TEAEs suspected of worsening retinopathy and a summary of the results of the follow-up dilated fundoscopic exam will be summarized by treatment and PT. The cases with follow-up fundoscopy during the course of the trial, based on clinical suspicion of worsening retinopathy that have either findings of de novo retinopathy or progression of retinopathy, and severe/serious adverse events from PTs defined in searching criteria in Appendix 1 (Section 6.1) will be considered AESIs and summarized.

3.14.3.11. Renal Safety

3.14.3.11.1. Acute Renal Events

Laboratory measures related to renal safety will be analyzed as specified for laboratory measurements in Section 3.14.4. Two shift tables examining renal function will be created. A min-to-min shift table of estimated glomerular filtration rate estimated by the CKD-EPI equation with units ml/min/1.73m^2, using categories (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/1.73m^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 300 mg/g, greater than 300 mg/g (respectively, these represent normal, microalbuminuria, and macroalbuminuria).

The AE database will be searched using SMQs of acute renal failure and chronic kidney disease to identify events consistent with acute renal events. The incidence of the resulting TEAEs will be summarized by treatment and PT. Detailed searching criteria can be found in Appendix 1 (Section 6.1). Severe/Serious acute renal events will be considered AESI.

3.14.3.11.2. Dehydration

The AE database will be searched using an SMQ of dehydration to identify events consistent with dehydration. Detailed searching criteria can be found in Appendix 1 (Section 6.1). Severe or serious dehydration events will be considered as AESIs.

3.14.4. Clinical Laboratory Evaluation

All laboratory data will be reported in the International System of Units. Selected laboratory measures will also be reported using conventional units. Limits from the performing laboratory 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.

3.14.5. Immunogenecity

TE ADA are defined as those with a titer 2-fold (1 dilution) greater than the minimum required dilution (1:10) 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 non-missing baseline ADA result, and at least 1 non-missing postbaseline ADA result.

The frequency and percentage of patients with preexisting ADA, with TE ADA, with neutralizing TE ADA, and with cross-reactive TE ADA to tirzepatide will be tabulated by dose, 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 reactions by TE ADA status will be tabulated if warranted by the data.

A listing may be provided of all immunogenicity assessments for those patients who at any time had TE ADA present. This includes the tirzepatide concentration from a simultaneous PK sample, and the clinical interpretation result. A listing may be provided for all participants who had ADAs present at any time (including baseline) or had any hypersensitivity or injection site reaction TEAEs.

Depending on the number of patients with TE ADA, selected efficacy and safety subgroup analyses by TE ADA categories may be performed if deemed necessary. TE ADAs that are associated with AEs of either severe/serious hypersensitivity or severe/serious injection site reaction will be classified as AESI.

3.14.6. Vital Signs

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

An MMRM using REML will be used to fit the changes from baseline in vital signs at all scheduled postbaseline visits. The model will include baseline T2DM status (Yes/No), the region (Japan, US including Mexico, and Europe including Israel), treatment group, visit, and

treatment-by-visit interaction as fixed effects, and baseline value of the dependent variable as a covariate. To model the covariance structure within patients, the unstructured covariance matrix will be used.

Counts and percentages of patients with abnormal sitting systolic BP, sitting diastolic BP, and pulse will be presented by treatment. The criteria for identifying patients with treatment-emergent vital sign abnormalities are stated in Table GPHR.3.3.

Table GPHR.3.3. Categorical Criteria for Abnormal 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	≥130 and increase from baseline ≥20
Diastolic BP (mm Hg) (supine or sitting - forearm at heart level)	≤50 and decrease from baseline ≥10	≥90 and increase from baseline ≥10
Pulse (bpm) (supine or sitting)	<50 and decrease from baseline ≥15	>100 and increase from baseline ≥15

Abbreviation: BP = blood pressure.

3.14.7. Electrocardiograms

For ECG parameters that collect in triplicates, the average from the 3 measurements for the same parameter at the same visit will be calculated and used for all the subsequent summaries and analyses.

Summary statistics by treatment and by nominal visit will be provided for ECG parameters (heart rate, pulse rate, QRS, QT, and QT corrected using Fredericia's correction factor [QTcF = QT / RR0.333] When the QRS is prolonged (e.g., a complete bundle branch block), QT and QTc 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: 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 GPHR.3.4.

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 GPHR.3.4
- QT greater than 500 msec
- QTcF greater than 500 msec, and

- treatment-emergent increase from the maximum baseline in QTcF interval of greater than 30 msec, 60 msec, or 75 msec at any time will be summarized. Maximum baseline will be the maximum nonmissing observation in the baseline period. The maximum value during the treatment period will be analyzed. Scheduled and unscheduled measurements will be included.

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

Table GPHR.3.4. 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
QTcF (msec)	<330	<340	>450	>470

Abbreviations: ECG = electrocardiogram; PR = pulse rate; QTcF = Fridericia's corrected QT interval.

3.15. Subgroup Analyses

Efficacy subgroup analyses will be guided by the efficacy estimand. Subgroups with few subjects may be excluded from subgroup analyses when appropriate. Subgroup analyses will include, but are not limited to:

- Primary and secondary endpoints by baseline T2DM status.
- Primary endpoint by baseline NAS score, using median NAS score to divide into 2 subgroups. In the case that baseline NAS score median is 4, NAS score 4, and NAS score greater than 4 will be the 2 subgroups.
- Primary endpoint by baseline liver fat (MRI-PDFF) category, using median to divide into 2 subgroups.
- Primary endpoint by baseline cT1 category, using median to divide into 2 subgroups.
- Primary endpoint by baseline serum NASH biomarkers (ALT, K-18, NIS4), using median (or cutoffs based on literature) to divide into 2 subgroups for each biomarker.
- Primary endpoint by categories of weight loss, using <5% and ≥5% as categories and then <10% and ≥10% as categories.
- Secondary endpoints of fibrosis regression and fibrosis progression by baseline fibrosis stage category (2 versus 3).
- Secondary endpoint of decreasing NAS by ≥2 points by baseline NAS.

- Secondary endpoints of fibrosis regression and fibrosis progression by baseline categories of serum fibrosis biomarkers (ELF and Pro-C3). For ELF, subgroups <9.8 and ≥9.8 (threshold for advanced fibrosis) will be analyzed (EASL 2021). For Pro-C3, subgroups <13.4 versus ≥13.4 and <20.2 versus ≥20.2 ng/ml (Nielsen et al. 2021) will be analyzed. If these subgroups for ELF and Pro-C3 are very imbalanced, the medians will be used to create the subgroups.

3.16. Important Protocol Violations

Important protocol deviations are identified in the Trial Issues Management Plan. A listing and a summary of important protocol deviations by treatment will be provided.

3.17. Interim Analyses and Data Monitoring

There will be a futility interim analysis based on relative liver fat reduction based on MRI-PDFF measurements. It will be done based on data from approximately 80 participants being on study treatment for 26 weeks. An internal AC will be formed to review unblinded interim analysis.

Detailed information regarding futility analysis and safety reviews including the statistical reports to be reviewed by the AC will be described in the program AC Charter and a separate interim analysis SAP.

4. Unblinding Plan

Details of the blinding and unblinding will be provided in Blinding and Unblinding Plan document for Study GPHR.

5. Novel Coronavirus (COVID-19) Impact

This section lists the potential statistical analyses that may be performed to assess the impact of COVID-19 pandemic when appropriate.

5.1. General Consideration

Percentage and count of randomized participants who followed the COVID-19 mitigation plan may be summarized by treatment group. This includes, but not limited to, participants rescreened, procedures conducted via remote visit or mobile home health visit, visits occurred using the extended visit windows, alternative way of investigator product shipment/dispensing, use of a local laboratory, etc.

Percentage and count of randomized participants who completely missed at least 1 study visit due to COVID-19 pandemic may also be summarized.

5.2. Exposure

A listing of randomized participants who had study drug temporary interruption due to COVID-19 pandemic may be provided.

5.3. Protocol Deviation

Percentage and count of randomized participants having important protocol deviation related to COVID-19 pandemic may be summarized by treatment.

Percentage and count of randomized participants with protocol deviation related to COVID-19 pandemic may also be summarized by treatment.

A listing of all randomized participants who had important protocol deviation due to COVID-19 pandemic may be provided.

5.4. Patient Disposition

A summary table for all randomized participants that discontinue study or study treatment due to COVID-19 pandemic may be provided by treatment.

A listing of randomized participants who discontinued the study or study treatment due to COVID-19 infection may be provided.

5.5. Adverse Events

A listing of randomized participants who had COVID-19 infection, including death due to COVID-19, during the postrandomization period may be provided.

5.6. Local Laboratory

If local laboratory data was brought into the Lilly database, additional details around how to combine with central laboratory data and analyze it may be provided.

5.7. Missing Data Due to COVID-19

For the primary endpoints and key secondary endpoints, missing data due to COVID-19 may be handled as described in Section 3.11.1.3.

6. Supporting Documentation

6.1. Appendix 1: Searching Criteria for Adverse Events of Special Interest

The AESI analyses are detailed in Section 3.14.3. The search criteria for each AESI are stored in CLUWE: \\statsclstr\lillyce\prd\ly3298176\common\AESI_Lab\Search criteria AESIs_GPHR.xlsx.

7. References

[ADA] American Diabetes Association Professional Practice Committee; Draznin B, Aroda VR, Bakris G, et al. 6. Glycemic targets: standards of medical care in diabetes-2022. *Diabetes Care*. 2022;1;45(suppl 1):S83-S96. https://doi.org/10.2337/dc22-S006

[EASL] European Association for the Study of the Liver. EASL clinical practice guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. *J Hepatol*. 2021;75(3):659-689. https://doi.org/10.1016/j.jhep.2021.05.025

Kleiner DE, Brunt EM, Van Natta M, et al.; Nonalcoholic Steatohepatitis Clinical Research Network. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005;41(6):1313-1321. https://doi.org/10.1002/hep.20701

Nielsen MJ, Leeming DJ, Goodman Z, et al. Comparison of ADAPT, FIB-4 and APRI as non-invasive predictors of liver fibrosis and NASH within the CENTAUR screening population. *J Hepatol*. 2021;75(6):1292-1300. https://doi.org/10.1016/j.jhep.2021.08.016

Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons Inc.; 1987.

Signature Page for VV-CLIN-081539 v1.0

Approval	PPD
Approval	

Signature Page for VV-CLIN-081539 v1.0