Statistical Analysis Plan I4L-GH-ABET

A Prospective, Randomized, Open-Label Comparison of a Long-Acting Basal Insulin Analog LY2963016 to Lantus® in Adult Chinese Patients with Type 2 Diabetes Mellitus

NCT03338010

Approval Date: 21-May-2020

1. Statistical Analysis Plan:

I4L-GH-ABET

A Prospective, Randomized, Open-Label Comparison of a Long-Acting Basal Insulin Analog LY2963016 to Lantus® in Adult Chinese Patients with Type 2 Diabetes Mellitus

LY2963016

Phase 3, randomized, multicenter, 2-arm, active-control, open-label, parallel, 24-week treatment study with a 4-week post-treatment follow-up to compare LY2963016 and Lantus® in adult Chinese patients with type 2 diabetes mellitus.

Eli Lilly and Company
Indianapolis, Indiana USA 46285
Protocol I4L-GH-ABET
Phase 3

Approval Date: 21-May-2020 GMT

2. Table of Contents

Section **Page**

1. Statistical Analysis Plan: 1
2. Table of Contents 2
3. Revision History 5
4. Study Objectives 6
   4.1. Primary Objective 6
   4.2. Secondary Objectives 6
5. A Priori Statistical Methods 7
   5.1. Determination of Sample Size 7
   5.2. Statistical and Analytical Plans 7
      5.2.1. General Considerations 7
      5.2.2. Patient Disposition 9
      5.2.3. Patient Characteristics 9
      5.2.4. Concomitant Therapy 9
      5.2.5. Treatment Compliance 10
      5.2.6. Protocol Violations 10
      5.2.7. Primary Efficacy Outcome and Methodology 10
      5.2.8. Secondary Efficacy Outcome and Methodology 11
      5.2.9. Pharmacokinetic/Pharmacodynamic Analyses 12
      5.2.10. Health Outcome/Quality of Life Analyses 12
         5.2.10.1. Insulin Treatment Satisfaction Questionnaire 12
      5.2.11. Safety Analyses 13
         5.2.11.1. Adverse Events 13
            5.2.11.1.1. Special Topic Assessment of Allergic Events 14
            5.2.11.1.2. Assessment of Injection Site Adverse Events 14
         5.2.11.2. Hypoglycemic Events 14
         5.2.11.3. Laboratory Measures – Chemistry/Hematology Panel 16
         5.2.11.4. Laboratory Measures – Insulin Antibodies 16
         5.2.11.5. Vital Signs 18
         5.2.11.6. Other Safety Measures 19
      5.2.12. Subgroup Analyses 20
      5.2.13. Interim Analyses 22
      5.2.14. Exploratory Analysis 22
      5.2.15. Required Analyses for the Clinical Trial Registry (CTR) 22
6. Unblinding Plan 23

7. References 24
8. Appendices 25

Table of Contents

Appendix **Page**

Appendix 1. List of Allergic Reaction Terms....26

Appendix 2. Estimation of Creatinine Clearance Using the Cockroft-Gault and MDRD for the Approximation of Glomerular Filtration Rate (GFR)....28

3. Revision History

SAP Version 1 was approved prior to the First Randomized Patient.

Version 2 of the SAP will be approved prior to the first unblinding of the clinical team and prior to the database lock.

The following changes have been incorporated:

- The imputation rule for missing scores in ITSQ is updated in section 5.2.10.1.
- The analysis of nonnocturnal hypoglycemia and pseudo hypoglycemia is deleted.
- Subgroup analysis of entry age (<75, ≥75) and entry HbA_{1c} levels (<7%, ≥7%) are deleted. BMI cut point changed from 28, 30 to 24, 28
- One more rule added to exclude subjects from Per Protocol Set: Missing HbA1c at baseline and/or at primary endpoint (week 24; only for subjects who should have V18)
- The definition of TEAR is updated.
- The shift tables of total inuslin antibody and cross reactive insulin antibody are deleted.
- The analysis concerning to the following hypoglycemic cagetories were deleted:
  - Severe hypoglycemia (BG level ⩽ 70 mg/dL[3.9 mmol/L])
  - Severe hypoglycemia (not biochemically confirmed—BG missing)
  - Severe hypoglycemia (not biochemically confirmed—BG not aligned with severe symptoms)
  - Nonnocturnal hypoglycemia
  - Pseudo-hypoglycemia

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to test the hypothesis that LY2963016 administered QD is noninferior to Lantus® administered QD by a margin of 0.40%, as measured by change in hemoglobin A1c (HbA1c) from baseline to 24 weeks, when used in combination with OAMs.

4.2. Secondary Objectives

The secondary objectives of the study are as follows:

- To test the hypothesis that Lantus® is noninferior to LY2963016 (QD), as measured by change in HbA1c from baseline to 24 weeks, when used in combination with OAMs. (this secondary objective is tested with a gated approach).
- To compare safety of LY2963016 relative to Lantus® (proportion of patients with detectable anti-glargine antibodies, hypoglycemia, and injection site reaction) when used in combination with OAMs.
- To compare change in HbA_{1c} at 4, 8, 12, 16, and 20 weeks between LY2963016 and Lantus® when used in combination with OAMs.
- To compare 7-point self-monitored blood glucose (SMBG) profiles (as plasma equivalent values) at 0, 2, 6, 12, and 24 weeks between LY2963016 and Lantus® when used in combination with OAMs.
- To compare percentage of patients with HbA1c <7% and percentage of patients with HbA1c ≤6.5% at 4, 8, 12, 16, 20, and 24 weeks between LY2963016 and Lantus® when used in combination with OAMs.
- To compare LY2963016 to Lantus® when used in combination with OAMs with regard to the following measures:
    - intrapatient blood glucose (BG) variability
    - basal insulin dose
    - weight change
- To compare LY2963016 relative to Lantus® for patient-reported outcomes (PRO) as measured by responses to the Insulin Treatment Satisfaction Questionnaire (ITSQ).

5. A Priori Statistical Methods

5.1. Determination of Sample Size

Based on the primary objective, to show noninferiority of LY2963016 to Lantus® at the 0.40% noninferiority margin (NIM), 450 completers in total with a ratio 2:1 (LY2963016 versus Lantus®) are needed at 24 weeks. This calculation assumes no treatment difference in HbA1c between LY2963016 and Lantus®, common standard deviation (SD) of 1.3% for change from baseline in HbA1c, 0.05 two-sided significance level, and over 85% power. Assuming a 15% dropout rate at 24 weeks, the required number of randomized patients is 530 in total (353 for LY2963016 and 177 for Lantus®).

5.2. Statistical and Analytical Plans

5.2.1. General Considerations

All data will be entered, verified, and archived by a contract research organization (CRO) external to Lilly and/or at Lilly. Data listings, summaries, and analyses will be performed by a CRO and/or by Lilly under the guidance and approval of statisticians at Lilly. Statistical analysis of this study will be the responsibility of Lilly.

Any change to the data-analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the change, will be described in the statistical analysis plan and/or in the clinical study report (CSR). Additional exploratory analyses will be conducted as deemed appropriate.

The patient populations used in the study are described below:

1. All Patients Entered - all patients who signed ICF
2. All Randomized - all patients who were randomized to a treatment arm
3. Full Analysis Set (FAS) - based on the intent to treat (ITT) principle, all patients who were randomized and who have taken at least one dose of study medication will be included in this analysis set. Patients will be analyzed according to the treatments to which they were randomized.
4. Per-protocol (PP) - patients in the FAS/ITT population who also meet the criteria in Tiral Issue Management Plan (TIMP), and treatment group will be defined on the basis of the treatment the patients actually receive:
    a. have no violations of Inclusion/Exclusion Criteria specified in TIMP
    b. have not discontinued from the study prior to 24 weeks
    c. have not been off study medication for more than 14 consecutive days during the treatment period

d. have not received chronic (lasting longer than 14 consecutive days) systemic glucocorticoid therapy (excluding topical, intra-articular, intraocular, and inhaled preparations)

e. Missing HbA1c at baseline and/or at primary endpoint (week 24; only for subjects who should have V18)

Unless otherwise specified, listings will be prepared using all randomized patients. Efficacy and safety analyses will be conducted using the FAS population. Selected analyses will be conducted using the All Randomized population and the PP population.

Unless otherwise noted, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and confidence intervals (CIs) will be calculated as 2-sided 95% CIs. All tests of interactions between treatment groups and other factors will be conducted at a 2-sided alpha level of 0.05. No adjustments for multiplicity will be performed. See Section 5.2.7 for the gate-keeping strategies used for primary/secondary endpoints.

The baseline is Visit 2. If baseline data are missing, the last measurement taken prior to this visit will be used for the baseline measurement.

The last visit for the 24-week treatment study is Visit 18. If the Visit 18 measurement is missing, the last post-baseline value will be carried forward to create the 24-week endpoint value using the last-observation-carried-forward (LOCF) methodology. If there are no measurements after Visit 2 (date of randomization), the 24-week endpoint values will be considered missing. Efficacy analyses only be conducted for patients with both nonmissing baseline value and at least 1 nonmissing post-baseline value.

The LOCF methodology will be utilized in the endpoint analyses using analysis of covariance (ANCOVA) of HbA_{1c}, laboratory chemistry and hematology, as well as analysis of hypoglycemia (rate and incidence) and insulin antibodies.

Unless otherwise noted, the analysis of the continuous secondary efficacy and safety measurements (SMBG at each time point and summaries of SMBG, weight, dose, and vital signs) will use the same mixed model repeated measure (MMRM) methodology for the primary efficacy analyses with the baseline value of the response variable as a covariate with the FAS patient population. Continuous laboratory measures will be analyzed using an analysis of covariance (ANCOVA) model.

For categorical measures, Fisher's exact test will be used. To avoid computational problems, a maximum computation time of five minutes will be programmed into the analysis for Fisher's exact test. If it does not converge in that time, the Pearson's Chi-square test will be utilized.

Values for the 7-point self-monitored blood glucose (SMBG) blood glucose profiles will be averaged over the two 7-point SMBG profiles obtained during 2-week period prior to each office visit (as specified in Protocol Section 7.2.3). For the average blood glucose calculation for a specific time point, if only 1 of 2 days of data is collected, then the value of the 1 day will be used. If 2 days of data are collected, then the average of the 2 days will be used. If more than 2

days of data are collected then choose the 2 according to the ranking first by the number of non-missing blood glucose measurements and then by day closest to the visit day. The handling of missing data for the ITSQ questionnaire is specified in Section 5.2.10.

Visits for laboratory data will be handled by conventions in the Lilly diabetes white paper (Standardization of Derivations for Efficacy Lab Variables).

All analyses will be implemented using SAS Version 9.4® or higher.

5.2.2. Patient Disposition

Summary of analysis population for each treatment group and combined groups will be presented. A listing of the primary reason for patient discontinuation will be presented for all randomized patients. Summary of patients disposition of study and study treatment will be conducted for the all randomized populations, and its frequency counts and percentages will be presented for each treatment group and combined groups. Summary of number of patients entered, number of randomized patients, and number of patients discontinued study during treatment period will be summarized by site.

A listing of subjects with treatment assignement and stratification categories will be created for all randomized patients.

A listing of subjects with visit date of each visit will be presented. A second listing of dates with first patient visit (visit 1), randomization (visit 2), last treatment visit (visit 18 or ED visit) and safety follow-up visit (V801).

A listing of subjects indicating its population flag and assigned treatment group will be provided for all patients entered.

A listing of subjects discontinue from study or study treatment due to COVID-19 will be provided.

5.2.3. Patient Characteristics

The patient's age, sex, weight, height, BMI, or other demographic characteristics will be recorded.

Demographic and baseline characteristics will be summarized by treatment group for the FAS and PP populations. For continuous measures, summary statistics will include sample size, mean, median, maximum, minimum and SDs. The treatment groups will be compared using a 2-sample *t* test. For categorical measures, summary statistics will include sample size, frequency, and percentage. Analysis will use Fisher's exact test. To avoid computational problems, a maximum computation time of five minutes will be programmed into the analysis for Fisher's exact test. If it does not converge in that time, the Pearson's Chi-square test will be utilized.

5.2.4. Concomitant Therapy

Concomitant medications, including previous therapy for diabetes, will be summarized by different categories and treatment group using the FAS population. All conconmitant therapies

that originally mapped using the WHODRUG dictionary in the Clintrial database will be reported using preferred term.

Oral antihyperglycemic medications (OAM) will be classified into classes of drugs [alpha glucosidase inhibitors, dipeptidyl peptidases (DPP) IV inhibitors, meglitinide, metformin, sulfonylurea, Sodium-glucose co-transporter 2 (SGLT2) inhibitors, and thiazolidinedione (TZD)]. Summary tables will be provided by treatment showing the number of OAMs and the particular combination classes of OAMs for patients with 2 OAMs, 3 OAMs, and > 3 OAMs.

5.2.5. Treatment Compliance

No specific study data will be collected for analysis of treatment compliance.

5.2.6. Protocol Violations

A comprehensive listing of patients with important protocol deviations (IPD) or protocol violations that could potentially impact data interpretation, data integrity and patient safety across the I4L-GH-ABET study will be provided. A listing of protocol deviations (PD) due to coronavirus disease 2019 (COVID-19) will be provided. A summary of IPD and IPD due to COVID-19 by treatment group and overall will also be provided.

Important protocol deviations will be identified from the clinical database and from site monitoring. Categories of IPDs will be documented in the "ABET trial issue management plan" that will contain detailed criteria used to identify IPDs. Detailed programming specifications for IPDs (ADaM PD dataset specification) will be stored in CLUWE prior to database lock.

5.2.7. Primary Efficacy Outcome and Methodology

The primary efficacy outcome will be the change in HbA1c level from baseline to 24 weeks. The primary analysis will be a likelihood-based, MMRM approach, treating the data as missing at random (MAR) for the FAS population. The MMRM model will evaluate the change from baseline to each post-baseline visit in HbA1c level as the dependent variable with treatment (LY2963016, Lantus®), entry use of insulin scretagogue (SU, meglitinide, neither), visit, and interaction between visit and treatment as fixed effects; the baseline value of HbA_{1c} as a covariate; and a random effect for patient. An unstructured covariance structure will be used to model the within-patient errors. Significance tests will be based on LS means and Type III tests. If this analysis fails to converge, the following covariance structures will be tested in order:

- Toeplitz with heterogeneity
- autoregressive with heterogeneity
- compound symmetry with heterogeneous variances
- Toeplitz
- autoregressive
- compound symmetry without heterogeneous variances

The first covariance structure that converges will be used. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom.

The primary treatment comparison is to compare LY2963016 versus Lantus® at the NIM of +0.4%. If the upper limit of the 95% CI on the change from baseline to 24-week endpoint HbA_{1c} for LY2963016 versus Lantus® is below +0.4%, then LY2963016 will be declared noninferior to Lantus®. The LSMean and standard error derived from the MMRM model for each treatment will be used to test noninferiority. Type III sums of squares will be used to make the treatment comparisons.

- If the +0.4% NIM is met, a key secondary treatment comparison is to compare Lantus® versus LY2963016 at the NIM of -0.4%. If the lower limit of the 95% CI on the change in HbA1c from baseline to the 24-week endpoint for LY2963016 versus Lantus® is above -0.4%, then Lantus® will be declared noninferior to LY2963016. The LSMean and standard error derived from the MMRM model for each treatment will be used to test noninferiority. This gate-keeping procedure controls the family-wise Type 1 error rate at a 1-sided 0.025 level.
- If LY2963016 is declared noninferior to Lantus® in the primary treatment comparison, and Lantus® is declared noninferior to LY2963016 in this secondary treatment comparison, then LY2963016 will be considered to have equivalent efficacy as Lantus®.

A first secondary analysis of the primary efficacy outcome will use the same MMRM model described above with the PP patient population. Significance tests will be based on LSMeans using the Type III sum of squares, and testing for noninferiority will occur as described above.

A second secondary analysis of the primary efficacy outcome will use an ANCOVA model with FAS population. The ANCOVA model will include treatment and entry use of insulin screatagogue (SU, meglitinide, neither) as fixed effects and the baseline value of HbA1c as covariate. The ANCOVA model will be carried out using the OM option in SAS GLM procedure. If the 24-week HbA1c value is missing, the last post-baseline value will be carried forward and used in the analysis. This creates the 24-week endpoint value for HbA1c using the LOCF methodology. If there are no HbA1c data after the date of randomization, the endpoint will be considered missing and the patient will not be included in the analysis.

The analyses of the primary efficacy outcome will only be conducted for patients with both nonmissing baseline value and at least 1 nonmissing post-baseline value.

5.2.8. Secondary Efficacy Outcome and Methodology

The continuous secondary efficacy outcomes include:

- Actual and change in HbA1c from baseline to 4, 8, 12, 16, and 20 weeks or LOCF
- 7-point SMBG measurements as listed in the Study Schedule (see ABET Study Protocol Attachment 1)
    - premeal for each meal

    - postmeal for each meal
    - bedtime
- summaries of 7-point SMBG (actual and change from baseline)
    - daily mean BG level (average across all 7 time points)
    - daily mean premeal BG level (before breakfast, lunch, dinner)
    - daily mean postprandial BG level (breakfast, lunch and bedtime)
- intrapatient variability, as measured by the SD of the 7-point SMBG
- actual and change from baseline in weight and BMI
- actual and change from baseline in basal insulin does (U/day and U/kg/day)
- patient-reported outcomes as reflected in responses to ITSQ

The analysis of the change from baseline of continuous secondary efficacy variables will be performed using the same MMRM model for the primary efficacy analysis with treatment (LY2963016, Lantus®), entry use of insulin scretagogue (SU, meglitinide, neither), visit, and interaction between visit and treatment as fixed effects; the baseline value of the response variable and the baseline HbA_{1c} value as covariates; and a random effect for patient.

The proportion of patients achieving HbA1c target values (HbA1c level <7.0% and ≤6.5%) at any point during the study (Weeks 4, 8, 12, 16, 20, and 24 and 24-week endpoint [LOCF]) will be analyzed using Fisher's exact test or Pearson's chi-square test.

A plot of the actual and change from baseline in weight (kg) will be presented.

5.2.9. Pharmacokinetic/Pharmacodynamic Analyses

Not applicable.

5.2.10. Health Outcome/Quality of Life Analyses

The Insulin Treatment Satisfaction Questionnaire (ITSQ) will be completed at Weeks 4 (Visit 6), 12 (Visit 12) and 24 (Visit 18) or ED.

The following will describe the analyses details for ITSQ.

5.2.10.1. Insulin Treatment Satisfaction Questionnaire

The ITSQ is a validated instrument containing 22 items that assesses treatment satisfaction completed at weeks 4 (visit 6), 12 (visit 10), and 24 (visit 18) or early discontinuation for persons with diabetes who are taking insulin. Items are measured on a 7-point scale in which lower scores reflect better outcomes. In addition to an overall score, the items that make up the 5 domains of satisfaction are categorized as:

- Inconvenience of Regimen (IR - 5 items)
- Lifestyle Flexibility (LF - 3 items)

- Glycemic Control (GC - 3 items)
- Hypoglycemic Control (HC - 5 items)
- Insulin Delivery Device (DD - 6 items).

All individual patient raw domain scores will be calculated as the mean of the individual item scores in the domain. If an item score is missing for a patient and <20% of the items within the domain are missing for that patient, then the mean of the items in the domain will be imputed for the missing item score(s). Individual patient with equal or more than 20% missing items in a domain, the domain score will be set to missing.

The individual transformed domain scores are then calculated as follows: use the following formula to transform the raw domain scores on a scale from 0-100 where higher scores indicate better treatment satisfaction:

Transformed domain score= 100*[(7- raw domain score)/6]

The individual raw and transformed overall scores are then calculated as follows: first, calculate the raw overall score as the mean of the raw domain scores for that patient. If any of the domain scores is missing, then set the raw overall score to missing. Then, use the following formula to transform the raw overall score on a scale from 0-100 where a higher score indicates better treatment satisfaction:

Transformed overall score= 100*[(7- raw overall score)/6]

Only the transformed domain and overall scores will be analyzed using the MMRM model for the FAS population. The MMRM model will have treatment (LY2963016, Lantus®), entry use of insulin scretagogue (SU, meglitinide, neither), visit, and interaction between visit and treatment as fixed effects; the baseline HbA_{1c} value as covariate; and a random effect for patient.

5.2.11. Safety Analyses

5.2.11.1. Adverse Events

Analyses of adverse events will include all data collected during the course of the entire study including the follow up visit, regardless of IP use.

Adverse events will be listed by patient, system organ class, Medical Dictionary for Regulatory Activities® (MedDRA) preferred term, severity, and relationship to the study disease, drug, device, or procedure for all patients. Adverse events (including injection site reactions, allergic events, and neoplasms) will be summarized as TEAEs for the FAS. Treatment-emergent adverse events are defined as events that are newly reported after first study treatment following randomization or reported to have worsened in severity from baseline. The proportion of patients experiencing each TEAE will be presented by preferred term, system organ class (SOC), and treatment group. The proportion of patients experiencing each TEAE that is assessed as possibly related to the study disease, drug, device, or procedures will also be summarized. The number of patients and proportion will be presented and compared by treatment using Fisher's

exact test or Pearson's chi-square test for the FAS population. TEAE will be summarized by preferred term within SOC and by preferred term by decreasing frequency.

All SAEs will be listed by patient and summarized by treatment as counts and percentages. If a sufficient number of SAEs are reported, the proportion of patients with SAEs between treatment groups will be compared using Fisher's exact test. Similar analyses will be performed for discontinuations due to AEs.

Summary table of historical conditions and pre-existing condition will be listed by preferred term and ordered by decreasing order of the total group. Historical conditions are conditions that end prior to inform consent and preexisting conditions are conditions that are still ongoing at inform consent or the conditions that started after inform consent and before first does.

A listing of subjects discontinue IP or study due to advere event including death will be presented.

5.2.11.1.1. Special Topic Assessment of Allergic Events

The special topic assessment of allergic events will be performed by an initial blinded review of preferred terms (PTs) by SOC in order to identify all possible cases of allergic events, followed by a comparison between treatment arms. The goal of the initial blinded review, which will be carried out by the safety physician, is to identify all reported allergic reactions. As an initial reference for this blinded review, Lilly will consider the list of PTs shown in Appendix 1. Justification in including or excluding events based on medical judgment and other supportive information will be provided when applicable. All allergic events will be listed by patient and summarized by treatment as counts and percentages. The proportion of patients with allergic reactions between treatment groups will be compared using Fisher's exact test or Pearson's chi-square test for the FAS.

5.2.11.1.2. Assessment of Injection Site Adverse Events

Whenever an injection site adverse event occurs, there is an evaluation of the pain, pruritus, and rash associated with the injection as well as of the characteristics of the injection site (abscess, nodule, lipoatrophy, lipohypertrophy, or induration). The proportion of patients experiencing treatment emergent injection site adverse events will be summarized and analyzed by treatment group. Additional analyses will be done as deemed appropriate.

5.2.11.2. Hypoglycemic Events

A **hypoglycemic episode** is that at any time a patient feels as he/she is experiencing a sign or symptom that is associated with hypoglycemia or has a BG level of ≤70 mg/dL (≤3.9 mmol/L), even if it was not associated with signs, symptoms, or treatment.

Severe hypoglycemia: an event requiring assistance of another person to actively administer carbohydrates or glucagon or take other corrective actions. Plasma glucose concentrations may not be available during an event, but neurologic recovery following the return of plasma glucose to normal levels is considered sufficient evidence that the event was induced by low plasma glucose concentration.

Nocturnal hypoglycemia: any total hypoglycemic event that occurs after bedtime and prior to the first meal upon waking (eg, breakfast).

Documented symptomatic hypoglycemia: an event during which typical symptoms of hypoglycemia are accompanied by a measured plasma glucose concentration ≤70 mg/dL (≤3.9 mmol/L).

Asymptomatic hypoglycemia: an event not accompanied by typical symptoms of hypoglycemia but with a measured plasma glucose concentration ≤70 mg/dL (≤3.9 mmol/L).

Probable symptomatic hypoglycemia: an event during which symptoms typical of hypoglycemia are not accompanied by a plasma glucose determination but that was presumably caused by a plasma glucose concentration ≤70 mg/dL (≤3.9 mmol/L).

Besides the hypoglycemic categories listed above, total hypoglycemic event will be defined as follows:

Total hypoglycemic event: any event which is either documented symptomatic hypoglycemia, asymptomatic hypoglycemia, probable symptomatic hypoglycemia or severe hypoglycemia.

For each category of hypoglycemia event, the incidence, the number of hypoglycemic events per patient, the rate of hypoglycemic events per year (that is, the number of hypoglycemic episodes per patient per patient year [365.25 days]) will be calculated. These measures will be summarized at baseline, titration, maintenance, and overall study periods and at endpoint.

The proportion of patients with at least 1 hypoglycemic event (total, severe, nocturnal, and others) or incidence during the study will be summarized (counts and percentages) and analyzed using Fisher's exact test or the Pearson's chi-square test for the FAS population.

Logistic regression will be used as a sensitivity analysis of the incidence of hypoglycemia. The model will have presence or absence of hypoglycemia as the dependent or response variable and treatment, baseline HbA_{1c}, and entry use of insulin scretagogue (SU, meglitinide, neither) as the independent terms in the model.

The rate of hypoglycemic episodes per year (total, severe, nocturnal, and others) will be analyzed at baseline, titration, maintenance, and overall study periods and at endpoint using the Wilcoxon test. In addition, the hypoglycemia rates will also be analyzed using a negative binomial model for the FAS population with terms for treatment, baseline HbA_{1c}, and entry use of insulin scretagogue (SU, meglitinide, neither).

In addition, the total number of patients with at least 1 hypoglycemic episode divided by the total extent of exposure in patient-years will be calculated for the overall study period and summarized descriptively for each treatment group for total, severe, nocturnal, documented symptomatic, asymptomatic hypoglycemia and probable sympotamic hypoglycemia. Individual patient listing of hypoglycemic events by visit will be presented for the FAS population.

All the above analysis will be repeated with the threshold of **Clinically significant hypoglycemia**: <54 mg/dL (3.0 mmol/L) for hypoglycemia categories of total hypoglycemia,

nocturnal hypoglycemia, documented symptomatic hypoglycemia and asymptomatic hypoglycemia.

5.2.11.3. Laboratory Measures – Chemistry/Hematology Panel

Continuous measures and their change from baseline to 24-week endpoint in the chemistry and hematology panels for the FAS population will be summarized using descriptive statistics at baseline and at 24-week endpoint for the FAS population.

The continuous measures and change from baseline values to 24 weeks will be analyzed using the ANCOVA model with treatment, entry use of insulin secretagogues (SU, meglitinide, neither) as fixed effects and the baseline value of HbA1c and the baseline of the response variable as covariates.

For each chemistry and hematology analyte, the number and percent of patients with treatment emergent high (not in high range at baseline, greater than the upper limit at Week 24/Early Discontinuation) and treatment emergent low (not in low range at baseline, less than the lower limit at Week 24/Early Discontinuation) will be presented and analyzed using Fisher`s exact test or Pearson's chi-square test. Unscheduled safety lab results will be included in the analysis.

5.2.11.4. Laboratory Measures – Insulin Antibodies

At each visit (baseline, visits 4, 6, 12, and 24-week (last observation carried forward [LOCF])), total insulin antibody status and its level of total percent binding will be collected. The lower limit of detection of total insulin antibodies is 1.19%. If antibodies are detected then cross reactive antibody status and its level of cross-reactive percent binding will be recorded. The lower limit of detection of cross-reactive antibodies is 1%.

Descriptive and inferential analyses will be performed for the set of FAS patients with a valid antibody testing (detected or ND) at baseline and at least one post baseline visit.

The proportion of patients with detected insulin antibodies will be summarized as counts and percentages at baseline, at each visit, at the 24-week endpoint (LOCF), and overall for the 24-week treatment period. At each of these time points, the proportion of patients with detected antibodies will be compared between treatment groups using Fisher's exact test. These analyses will be performed for patients with antibody levels detectable at baseline and for patients without detectable levels at baseline) in the FAS population.

In addition, the following listing will be provided for patients with detectable insulin antibodies at any time during the study: a listing of level of total insulin antibody percent binding across visits sorted by treatment and maximum postbaseline percent binding in descending order. Similar listings sorted by baseline percent binding as well as by endpoint percent binding will be provided. These listings will be provided for the FAS population.

The level of detectable insulin antibodies (expressed as percent binding) will be summarized by descriptive statistics (mean, median, SD, Q1, inter-quartile range (IQR), Q3, standard error, minimum, and maximum) at baseline, each visit, and endpoint (LOCF). At each of these time points, the median level of percent binding will be compared between treatment groups using the

Wilcoxon rank sum test. Graphical displays of median percent binding and IQR at each visit and endpoint (LOCF) by treatment will be presented. These analyses will be performed for the FAS population, for patients with antibody levels detectable at baseline and without detectable levels at baseline.

All of the analyses above will be performed for the level of cross-reactive antibodies as well.

The treatment-emergent antibody response (denoted TEAR throughout this SAP) is based on the change from baseline to post-baseline in the anti-insulin antibody level (percent binding). TEAR can be sub-classified as either treatment-induced (not detected anti-insulin antibody at baseline) or treatment-boosted (detected anti-insulin antibody at baseline):

- treatment-induced response: change from not detected anti-insulin at baseline to post-baseline detected anti-insulin;
- treatment-boosted response: change from detected anti-insulin at baseline to post-baseline detected anti-insulin antibody level (percent binding) at least 147% of the baseline value.

The number and proportion of patients who have a treatment-emergent antibody response (TEAR) will be summarized by treatment at each post-baseline visit, at the 24-week endpoint (LOCF), and overall for the 24-week treatment period, then analyzed using Fisher`s exact test or Pearson's chi-square test.

For patients with TEAR, a visit-wise listing of total and cross-reactive insulin antibody percent binding and clinical outcomes (HbA_{1c}, total hypoglycemia rate, basal insulin dose (U/Day, U/kg/Day) will be presented. The listing will be sorted by treatment and endpoint total insulin percent binding.

Incidence tables showing the number and percent of patients with initial and continuing TEAR at each visit will be presented. Similar analyses will be performed for the detection of total insulin antibodies and for the detection of cross reactive insulin antibodies.

Box plots of clinical outcomes (HbA_{1c}, total hypoglycemia rate, basal dose) will be generated by treatment and TEAR status.

In addition, the following patient listings of insulin antibodies (percent binding) will be provided for patients with TEAR: a listing of level of total insulin antibody percent binding across visit sorted by treatment and maximum postbaseline percent binding; a listing of level of total insulin antibody percent binding across visit sorted by treatment and endpoint percent binding. Similar listings will be presented for cross reactive percent binding.

In addition, the potential impact of total insulin antibody level and TEAR status on clinical response will be evaluated in several ways:

1. The relationship between the natural logarithm of the last observed insulin antibody levels (Ln[antibody level]) and selected clinical response variables (efficacy and safety measures (eg HbA_{1c}, total hypoglycemia rate, basal dose [U/day, U/kg/day],) will be evaluated using scatterplots and analyzed using ANCOVA on efficacy/safety

measure as the dependent variable with treatment, and entry use of insulin scretagogue (SU, meglitinide, neither) as fixed effects; baseline HbA_{1c} value, baseline response variable, Ln(antibody level) and treatment-by-Ln(antibody level) interaction as covariates for the FAS patients with detectable antibodies postbaseline and with valid antibody testing at baseline (detected, ND). A significant treatment-by-insulin antibody interaction may be indicative of a differential treatment effect necessitating further exploration to determine the nature of the interaction. Only patients with non-missing baseline value and at least one non-missing post-baseline value of the clinical response variable will be included in the analysis.

2. The relationship between TEAR and selected clinical response variables (efficacy and safety measures (e.g. HbA_{1c}, total hypoglycemia rate, basal dose [U/day, U/kg/day],) will be analyzed using ANCOVA on efficacy/safety measure as the dependent variable with treatment, TEAR (yes/no), entry use of insulin scretagogue (SU, meglitinide, neither), treatment-by-TEAR interaction as fixed effects, baseline response variable and baseline HbA_{1c} value as a covariate for the FAS patients with detectable antibodies postbaseline and with valid antibody testing at baseline (detected, ND). A significant treatment-by-TEAR interaction may be indicative of a differential treatment effect necessitating further exploration to determine the nature of the interaction. Only patients with non-missing baseline value and at least one non-missing post-baseline value of the clinical response variable will be included in the analysis.
3. The relationship between TEAR and overall incidence of categories of adverse events (TEAE, SAE, TEAE related to study drug, special topic allergic reactions, injection site reactions) will be assessed by showing the proportion of patients with an event for TEAR and non-TEAR patients by treatments for the FAS. For patients with and without TEAR, treatments will be compared using the Mantel-Haenszel test and the odds ratio and p-value from this test will be reported. The homogeneity of the odds ratios for TEAR and non-TEAR patients will be assessed using the Breslow-Day test.

The potential impact of cross-reactive antibody formation on clinical response similar to the analyses above will be done if there is sufficient number of patients with cross-reactive antibodies.

5.2.11.5. Vital Signs

Vital signs measures (systolic blood pressure, diastolic blood pressure, and heart rate) and their change from baseline will be summarized by descriptive statistics (mean, median, SD, standard error, minimum, and maximum) by visit for the FAS population.

In addition, the MMRM model will evaluate the change from baseline in vital sign measure as the dependent variable with treatment (LY2963016, Lantus®), entry use of insulin scretagogue (SU, meglitinide, neither), visit, and interaction between visit and treatment as fixed effects; the baseline value of the vital sign measure and the baseline HbA_{1c} valueas covariates; and a random

effect for patient. The LSMeans will be estimated using the observed margins (OM) option in SAS MIXED procedure.

The number and percentage of patients with treatment-emergent outlier vital sign measures during study treatment (overall) will be presented and analyzed using Fisher`s exact test or Pearson's chi-square test for systolic blood pressure, diastolic blood pressure, and pulse rate according to the categories as listed below:

- low systolic blood pressure (≤90 mmHg and a decrease from baseline ≥20 mmHg)
- high systolic blood pressure (≥140 mmHg and an increase from baseline ≥20 mmHg)
- low diastolic blood pressure (≤50 mmHg and a decrease from baseline ≥10 mmHg)
- high diastolic blood pressure (≥90 mmHg and an increase from baseline ≥10 mmHg)
- low heart rate (<50 bpm and a decrease from baseline ≥15 bpm)

high heart rate (>100 bpm and an increase from baseline ≥15 bpm)

5.2.11.6. Other Safety Measures

Exposure:

Exposure to each treatment during the treatment period of the study will be calculated for each patient and summarized by treatment group. Exposure will be calculated as the number of days from the date of first dose of the study drug (or if this information is missing, from the date of randomization) to the date of last treatment dose. Exposure will be expressed in days, months (30 days) and years (365.25 days).

Listings of Hepatic Disorders and Abnormal Liver Enzymes:

A listing will be provided for patients with treatment emergent adverse events in the following MedDRA SMQs for Hepatic Disorders (Broad and Narrow SMQ):

Narrow	Biliary disorders	20000118
Narrow	Drug-related hepatic disorders – comprehensive	20000006

Additionally listings will be provided for all patients who meet at least one of the following liver enzyme outlier criteria for at least one visit: ALT >3 ULN, AST >3 ULN, total bilirubin >2 ULN, or alkaline phosphatase >2 ULN at any visit. These listings will include all liver enzymes at all visits.

Listings Based on Renal Function:

Listings will be provided for patients with severe reduction in GFR or kidney failure at any visit according to the following criteria:

Severe reduction in eGFR (15-29 mL/min/1.73 m^2)
Kidney Failure (eGFR< 15 mL/min/1.73 m^2).

These listings will include all estimated GFR (eGFR) at all visits.

5.2.12. Subgroup Analyses

The consistency of the treatment effect for the change in HbA_{1c} level will be assessed within subgroups in the FAS population if there are sufficient numbers of patients in each subgroup (e.g. 20 patients per cell). The following subgroups will be analyzed:

- Entry HbA_{1c} levels (<8.5%, ≥8.5%)
- Entry BMI (<28, ≥28)
- Entry BMI (<24, ≥24)
- Entry age (<65, ≥65)
- Gender
- Entry use of insulin secretagogues (SU, meglitinide, neither)
- Renal function, as estimated by estimated glomerular filtration rate (EGFR) using the MDRD formula (see Appendix 2). The following EGFR categories will be used:
    - Normal or increased GFR: EGFR (≥90 mL/min/1.73 m^2)
    - Mild reduction in GFR: EGFR (60-89 mL/min/1.73 m^2)
    - Moderate reduction in GFR: EGFR (30-59 mL/min/1.73 m^2)
    - Severe reduction in GFR: EGFR (15-29 mL/min/1.73 m^2)
    - Kidney failure: EGFR (<15 mL/min/1.73 m^2).

The categories of severe reduction in GFR and kidney failure will be combined if there are fewer than 20 patients in either of these categories.

The change in HbA_{1c} from baseline to 24-week endpoint will be analyzed using MMRM with treatment (LY2963016, Lantus®), entry use of insulin scretagogue (SU, meglitinide, neither), visit, subgroup, subgroup-by-treatment interaction, subgroup-by-visit interaction, treatment-by-visit interaction, and treatment-by-visit-by-subgroup interaction as fixed-effects; the baseline value of HbA1c as a covariate, and a random effect for patient for the FAS. If the subgroup is one of the stratification variables, then the subgroup will only be included once in the model. A significant treatment-by-subgroup interaction ($p<.05$) may be indicative of a differential treatment effect across levels of the subgroup, necessitating further exploration of the nature of the interaction.

For each subgroup listed above, the change in weight from baseline to 24-week will be analyzed using the same methodology as the subgroup analysis of HbA_{1c}, except baseline weight will be added as another covariate.

Additional subgroup analyses will be carried out on selected safety outcomes:

a) total or nocturnal hypoglycemia at 24-week endpoint and for the overall 24-week period (rates and incidence)
b) treatment-emergent antibody response (TEAR) at 24-week endpoint and for the overall 24-week period
c) detectable insulin antibodies at 24-week endpoint and for the overall 24-week period
d) categories of adverse events:
    - treatment-emergent adverse events (TEAEs) (overall incidence and preferred terms with >5% incidence)
    - TEAEs related to study drug (overall incidence)
    - special topic assessment of allergic events (overall incidence)
    - injection site reactions (overall incidence)
    - serious adverse events (overall incidence)

For rates and incidence of hypoglycemia (total, nocturnal) and categories of adverse events, the same set of 8 subgroups used to analyze change in HbA1c and change in weight will be analyzed.

For each categorical safety outcome and subgroup listed above, the proportion of patients with an event will be compared between treatments for the FAS. Within each subgroup, treatments will be compared using the Mantel-Haenszel test and the odds ratio and p-value from this test will be reported. The homogeneity of the odds ratios across subgroups will be assessed using the Breslow-Day test.

For the 8 subgroups listed above, the rate of total, severe or nocturnal hypoglycemia per year at the 24-week endpoint and for the overall 24-week period will be analyzed. Within each subgroup, treatments will be compared using the negative binomial (NB) model with treatment as the only factor. The NB mean for each treatment and their ratio with corresponding 95% confidence interval will be presented. Additionally, within each subgroup the rates will be compared using the Wilcoxon test. The significance of the subgroup-by-treatment interaction will be evaluated using a negative binomial model with factors treatment, subgroup and treatment-by-subgroup interaction. The negative binomial model will be implemented using the SAS PROC GENMOD with natural logarithm of exposure time (in days) as an offset variable. The above subgroup analysis will be performed and p-values will be provided if there are at least a total of 20 events in the combined treatment groups.

For TEAR at 24 weeks endpoint and overall, subgroup analyses will be performed only for entry age (>65, ≤65) and renal function. For detectable antibodies at 24 weeks endpoint and overall, we will not do subgroup analyses.

Other subgroup analyses may be performed if deemed appropriate as exploratory analyses.

5.2.13. Interim Analyses

No interim analyses are planned for this study. If an unplanned interim analysis is deemed necessary, the appropriate Lilly medical director, or designee, will be consulted to determine whether it is necessary to amend the protocol.

5.2.14. Exploratory Analysis

Additional exploratory analyses will be conducted when deemed appropriate.

5.2.15. Required Analyses for the Clinical Trial Registry (CTR)

The study team will create the CTR adverse event dataset based on the CTRAESUMM ADaM standard. A member of the CTR team will be responsible for generating the standard CTR reports from the CTR adverse event dataset and for uploading the reports to the CTR site. The CTR adverse event dataset will include the following requirements:

- Both Serious Adverse Events and 'Other' Adverse Events will be summarized and analyzed:
    - An adverse event is considered 'Serious' whether or not it is a treatment emergent adverse event (TEAE)
    - An adverse event is considered in the 'Other' category if it is both a TEAE and is not serious
- Serious Adverse Events and 'Other' Adverse Events will be summarized: by treatment group and by MedDRA preferred term.
- For each Serious AE and 'Other' AE term, the following will be provided for each treatment group:
    - the number of participants at risk of an event
    - the number of participants who experienced each event team
    - the number of events experienced
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients in every treatment group may not be included if a 5% threshold is chosen (5% is the maximum threshold).
- AE reporting will be consistent with other document disclosures for example, the CSR, manuscripts, and so forth. A member of the CTR team will perform the quality checks to ensure that the AE reporting for the CTR is consistent.

6. Unblinding Plan

This is an open-label study in which investigators, patients, study site personnel, and study monitors will be aware of the treatment assignment. To minimize bias, review of summary data by the Lilly study team (i.e. CRP/CRS overseeing the global conduct of the study, statisticians, and statistical analysts) prior to the final database lock of the study (at the end of 24 weeks of treatment) will remain blinded to treatment assignment. Unblinding of an individual patient's study drug treatment assignment may occur in the course of consultation between the investigator and the study team (principally between the investigator and the CRP/CRS) or during review of SAEs. No systematic unblinding of study drug treatment assignments will be performed by the Lilly study team before the final database lock. Similar to a double-blind study, a minimum number of Lilly personnel will have access to the randomization table and treatment assignments before the final database lock.

7. References

[White Paper] White Paper on the Classification and Data Capture of Hypoglycemia Events. Lilly Unpublished Manuscript, version 2017.

8. Appendices

Appendix 1. List of Allergic Reaction Terms

Allergic bronchitis	Hypersensitivity	Pruritus generalized
Allergic colitis	Idiopathic urticaria	Rash
Allergic cough	Immediate post-injection reaction	Rash erythematous
Allergic cystitis	Injection site dermatitis	Rash follicular
Allergic keratitis	Injection site eczema	Rash generalized
Allergic oedema	Injection site erythema	Rash macular
Allergic otitis media	Injection site hypersensitivity	Rash maculo-papular
Allergic pharyngitis	Injection site induration	Rash maculovesicular
Allergic respiratory symptom	Injection site inflammation	Rash papular
Alveolitis allergic	Injection site macule	Rash pruritic
Anaphylactic reaction	Injection site nodule	Rash pustular
Anaphylactic shock	Injection site oedema	Rash vesicular
Anaphylactoid reaction	Injection site papule	Reaction to drug excipients
Anaphylactoid shock	Injection site photosensitivity reaction	Reaction to preservatives
Angioedema	Injection site pruritus	Reversible airways obstruction
Arthralgia	Injection site pustule	Scleral oedema
Arthritis	Injection site rash	Scleritis allergic
Arthritis allergic	Injection site reaction	Skin oedema
Asthma	Injection site recall reaction	Small bowel angioedema
Auricular swelling	Injection site streaking	Stevens-Johnson syndrome
Bronchial hyperreactivity	Injection site swelling	Stridor
Bronchial oedema	Injection site urticaria	Suffocation feeling
Bronchospasm	Injection site vesicles	Swelling face
Circumoral oedema	Joint effusion	Swollen tongue
Conjunctival oedema	Joint swelling	Throat tightness
Corneal oedema	Laryngeal obstruction	Tongue oedema
Dermatitis	Laryngeal oedema	Toxic epidermal necrolysis
Dermatitis allergic	Laryngitis allergic	Toxic skin eruption
Dermographism	Laryngotracheal oedema	Tracheal obstruction
Diffuse cutaneous mastocytosis	Lip oedema	Tracheal oedema
Drug eruption	Lip swelling	Type I hypersensitivity
Drug hypersensitivity	Local swelling	Type II hypersensitivity
Drug rash with eosinophilia and systemic symptoms	Localised oedema	Type III immune complex mediated reaction

Encephalopathy allergic	Nasal oedema	Type IV hypersensitivity reaction
Eosinophilic oesophagitis	Nephritis allergic	Urticaria
Epiglottic oedema	Oculorespiratory syndrome	Urticaria cholinergic
Erythema multiforme	Oedema mouth	Urticaria chronic
Erythema nodosum	Oedema mucosal	Urticaria contact
Eye oedema	Oesophageal oedema	Urticaria papular
Eye swelling	Orbital oedema	Urticaria physical
Eyelid oedema	Oropharyngeal swelling	Urticaria pigmentosa
Face oedema	Palatal oedema	Urticaria pressure
Gastrointestinal oedema	Periarthritis	Urticaria thermal
Gingival oedema	Periorbital oedema	Urticaria vesiculosa
Gingival swelling	Pharyngeal oedema	Urticaria vibratory
Haemorrhagic urticaria	Photosensitivity allergic reaction	Visceral oedema
	Photosensitivity reaction	Wheezing
	Pruritus	
	Pruritus allergic	

Appendix 2. Estimation of Creatinine Clearance Using the Cockroft-Gault and MDRD for the Approximation of Glomerular Filtration Rate (GFR)

The most common equations used in the United States are the Cockcroft-Gault and Modification of Diet in Renal Disease (MDRD) study equations. The IDMS MDRD study equation is increasingly utilized in the United States (Levey et al. 2006).

Cockcroft-Gault equation — The Cockcroft-Gault equation allows the creatinine clearance to be estimated from the serum creatinine in a patient with a stable serum creatinine:

Male eCrCl (mL/min) = (140 - age[years]) x (weight[kg]) / (sCr x 72)

Female eCrCl (mL/min) = (140 – age[years]) x (weight[kg) x 0.85 / (sCr x 72)

where sCr is the serum Creatinine in mg/dl.

This formula takes into account the increase in creatinine production with increasing weight, and the decline in creatinine production with age. For women, the formula requires multiplication by 0.85 to account for smaller muscle mass compared to men.

Modification of Diet in Renal Disease (MDRD) Equation

Estimate the patient's creatinine clearance using Modification of Diet in Renal Disease (MDRD) equation as an approximation of glomerular filtration rate (GFR). The MDRD equation is:

$$eGFR = 175 \times \text{standardized sCr}^{-1.154} \times \text{age}^{-0.203} \times 1.212 \text{ [if black]} \times 0.742 \text{ [if female]}$$

Renal function, as estimated by estimated glomerular filtration rate (eGFR) using the MDRD formula will be used to calculate the following eGFR categories (Stage 1 to Stage 5):

Stage 1: Normal or increased eGFR (>90 mL/min/1.73 m^2)
Stage 2: Mild reduction in eGFR (60-89 mL/min/1.73 m^2)
Stage 3: Moderate reduction in eGFR (30-59 mL/min/1.73 m^2)
Stage 4: Severe reduction in eGFR (15-29 mL/min/1.73 m^2)
Stage 5: Kidney Failure (eGFR <15 mL/min/1.73 m^2)

References:

Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron*. 1976;16(1):31-41.

Levey AS, Coresh J, Greene T, Stevens LA, Zhang Y (Lucy), Hendriksen S, Kusek JW, Van Lente F, for the Chronic Kidney Disease Epidemiology Collaboration. Using standardized serum creatinine values in the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate. *Ann Intern Med*. 2006;145:247-254.

Leo Document ID = bb061887-1507-46af-945a-7d33777e2b72

Approver: PPD

Approval Date & Time: 21-May-2020 10:15:22 GMT

Signature meaning: Approved