H9X-MC-GBDJ (d) REWIND Statistical Analysis Plan Version 4

Protocol H9X-MC-GBDJ (d) (REWIND) The Effect of Dulaglutide on Major Cardiovascular Events in Patients with Type 2 Diabetes: Researching Cardiovascular Events with a Weekly INcretin in Diabetes (REWIND)

NCT01394952

Approval Date: 06-Oct-2016

1. Statistical Analysis Plan:
The Effect of Dulaglutide on Major Cardiovascular Events in Patients with Type 2 Diabetes: Researching Cardiovascular Events with a Weekly INcretin in Diabetes (REWIND)

Confidential Information

The information contained in this Statistical Analysis Plan (SAP) is confidential and the information contained within it may not be reproduced or otherwise disseminated without the approval of Eli Lilly and Company or its subsidiaries. This document and its associated attachments or appendices are subject to United States Freedom of Information Act Exemption 4.

Dulaglutide (LY2189265)
Type 2 Diabetes Mellitus

This is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel study to assess the effects of dulaglutide (LY2189265) on cardiovascular outcomes in patients with type 2 diabetes who are drug naïve or who are on a stable anti-diabetic regimen.

Eli Lilly and Company
Protocol H9X-MC-GBDJ
Phase 3
Statistical Analysis Plan Version 1 electronically signed and approved by Lilly:
21 November 2011
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly:
07 August 2013
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly:
08 January 2015
Statistical Analysis Plan Version 4 electronically signed and approved by Lilly on date provided below

Approval Date: 06-Oct-2016 GMT

LY2189265

2.   Table of Contents

Section   **Page**

Table of Contents

Table   **Page**

Table of Contents

Revision History

The protocol for this study was approved on 02 March 2011. Protocol amendment (a) was approved on 03 June 2011. Protocol amendment (b) was approved on 27 February 2012. The SAP a Priori Analyses addresses the planned statistical analyses. Version 1 of this SAP was approved on 21 November 2011 prior to the first unblinding of treatment codes. It was updated in Version 2 following protocol amendment (b) and prior to unblinding of the sponsor or the steering committee.

Key additions and clarifications in Version 2 include:

- the addition of anti-vascular endothelial growth factor (VEGF) therapy to the secondary efficacy objectives of composite microvascular endpoint per amendment (b)

- the addition of an urgent heart failure (HF) visit to the secondary efficacy objective of HF requiring hospitalization per amendment (b)

    The replacement of race and/or ethnicity by race in the subgroup analyses due to the fact that information on ethnicity is required only of patients randomized in the United States.

Version 2 of this SAP was approved on 07 August 2013 prior to the first unblinding of treatment codes. It was updated in Version 3 and approved prior to unblinding of the sponsor or the steering committee. Key changes and additions in Version 3, as well as the rationale, include:

- The number of interim analyses was changed from 2 to 1 with the timing of the analysis set to detect a clear benefit if one was emerging. The alpha spending function was changed to the O'Brien-Fleming to also detect any emerging benefit. The timing of the interim analysis at 730 events would provide enough events to assess evidence of efficacy, harm, or futility, and to potentially stop the study if warranted.

- The decision rule at the interim analysis time point for stopping or continuing the trial was modified to include testing futility for superiority. This allows for the trial to be stopped if the regulatory requirements are met at the interim and if it is very unlikely that superiority would be demonstrated by the end of the trial

- To justify any potential claim of the secondary objectives, a graphical testing strategy was introduced to test the secondary hypotheses and to control the family-wise error rate across the testing of secondary objectives

- The heading text of Sections **4.2.2. Prespecified Safety Objectives**, and **5.17.1. Prespecified Safety Measures** were changed to **4.2.2. Prespecified Safety Objectives (Adverse Events of Special Interest) and 5.17.1. Prespecified Safety Measures (Adverse Events of Special Interest)** and additional items were listed to fulfill Lilly's post marketing commitment to Food and Drug Administration (FDA). The additional items included in the list are serious gastrointestinal (GI) events, specific categories of cancer including pancreatic cancer, and other thyroid carcinomas, immune mediated reactions, serious hepatic events, clinically significant supraventricular arrhythmias and cardiovascular conduction disorders and serious renal events

- Keeping only a mixed-effects model for repeated measures (MMRM) for the analysis of changes from baseline for continuous measures, instead of this model and an analysis of covariance (ANCOVA) model with multiple imputations since MMRM also (implicitly) accounts for missing values

- For clarity, the inclusion of more details on the analysis of continuous laboratory measurements

- For clarity, the reordering of the variables for subgroup analyses into 2 categories: pre-specified and exploratory

- For clarity, the description of the time-to-event analysis method in only 1 section that will be referenced in subsequent sections where it is mentioned

- The addition of a criterion for a baseline factor to be included as covariate in a covariate adjusted analysis

- Clarification of the definition of treatment exposure

- Addition of a variable of "Prior cardiovascular (CV) events" for multifactor-adjusted and potential subgroup analyses, with definition consistent with the protocol

- Additions of 2 sections for the analyses of fatal myocardial infarctions(MIs) and fatal strokes

- Addition of a section for the analysis of nonfatal CV events with the competing risk of death

- The addition of a sensitivity analysis of the primary endpoint excluding silent MI

- The addition of "Ratio of total cholesterol to high-density lipoprotein-cholesterol (HDL-C) (total cholesterol divided by HDL-C)" to the lipid parameters, and the reporting of medians for triglycerides

- Substitutions of MMRM models for ANCOVA for the analyses of continuous postbaseline measurements including Montreal Cognitive Assessment (MoCA), Digit Symbol Substitution Test (DSST) and International Index of Erectile Function (IIEF) scores

- For the analysis of severe hypoglycemia rates, the change from a Generalized Estimating Equation (GEE) model to the more appropriate generalized mixed effect model with a negative binomial distribution. Yearly rates will be analyzed instead of 30-day adjusted rates. Time to first severe hypoglycemia event will also be analyzed

- Removal of race from the subgroup analyses due to the fact that patients are allowed to check multiple subcategories of race therefore making these subcategories nonindependent

- The change to the content of Section 6.4 to match the content of "Unblinded Data" in the Independent Data Monitoring Committee (IDMC) charter

Version 3 of this SAP was approved on 08 January 2015. Version 4 is being approved prior to the interim analysis, and unblinding of the sponsor or the steering committee. Key changes and additions in Version 4, as well as the rationale, include:

- The increase of the total number of events for the final analysis to 1200 following the release of the LEADER trial results in order to increase the power for the primary and key secondary analyses

- The removal of the futility threshold for superiority from the interim decision rule based on the additional information from the LEADER trial results

- The clarification of the data cut-off date for the interim analysis

- The introduction of the Whitehead method to control for type I error across the interim analysis and the final analysis that will follow an interim IDMC decision to stop the trial for efficacy. These final analyses could be used by FDA for labeling discussions

- The inclusion of 2 sensitivity analyses to explore the impact of missing data on the MACE primary endpoint findings following feedback from the FDA about the importance of addressing missing data for other dulaglutide studies

- The update to the graphical testing scheme to control type I error in the analyses of the primary and key secondary objectives, following the release of the results from the LEADER and SUSTAIN-6 trials

- The update to the definition of prior CV event

- The update to the criteria for important protocol deviation

3.   Study Objectives

3.1.   Primary Objective

The primary objective is to test the hypothesis that a once-weekly injection of 1.5 mg dulaglutide reduces the occurrence of the composite primary endpoint of death from cardiovascular (CV) causes, nonfatal myocardial infarction (MI), or nonfatal stroke when added to the glucose-lowering regimen of patients with type 2 diabetes, compared to the addition of a once-weekly placebo injection.

3.2.   Secondary Objectives

3.2.1.   Efficacy Objectives

The secondary efficacy objectives are to assess the effects of add-on therapy with 1.5-mg dulaglutide compared to placebo on the occurrence of:

- the composite microvascular endpoint of diabetic retinopathy requiring laser therapy, vitrectomy, or anti-vascular endothelial growth factor (VEGF) therapy; development of clinical proteinuria, a 30% decline in estimated glomerular filtration rate (eGFR), or need for chronic renal replacement therapy

- hospitalization for unstable angina

- each component of the composite primary endpoint

- all-cause mortality

- heart failure (HF) requiring hospitalization or an urgent HF visit

3.2.2.   Prespecified Safety Objectives (Adverse Events of Special Interest)

The prespecified safety objectives are to assess the effects of add-on therapy with 1.5 mg dulaglutide compared to placebo on the incidence of:

- acute pancreatitis

- serious gastrointestinal (GI) events

- any cancer (excluding basal or squamous cell skin cancer) and specific categories of

  o pancreatic cancer

  o medullary thyroid carcinoma (MTC) and C-cell hyperplasia

  o thyroid carcinomas

- severe hypoglycemia

- immune mediated reactions including serious allergic and hypersensitivity reactions

- serious hepatic events

- clinically significant supraventricular arrhythmias and cardiovascular conduction disorders

- serious renal events

- discontinuation of study drug for any reason

3.2.3. Additional Objectives

The additional objectives are to assess the effects of 1.5 mg dulaglutide compared with placebo on the following:

- hemoglobin A1c (HbA1c) levels

- weight

- waist/hip ratio

- the composite endpoint of death from CV causes, nonfatal MI, nonfatal stroke, or hospitalization for unstable angina

- coronary, carotid, or peripheral revascularizations, individually and compositely

- any hospitalization

- cognitive function as measured by the Montreal Cognitive Assessment (MoCA) and the Digit Symbol Substitution Test (DSST)

- erectile function using the International Index of Erectile Function Questionnaire (IIEF)

- any fracture

- development of cholelithiasis

4.   A Priori Statistical Methods

REWIND will be a randomized, double-blind, placebo-controlled, international, multicenter, parallel-arm trial to determine whether the addition of the once weekly glucagon-like peptide-1 (GLP-1) analog dulaglutide to the diabetes regimen of patients with type 2 diabetes at high CV risk reduces major adverse CV and other serious outcomes. There will be a single-blind placebo run-in period to test a prospective participant's behavior and willingness to inject study drug on a weekly basis, given that patients will be expected to inject study therapy for 5 or more years.

There will be 1 interim analysis and 1 final analysis. The interim analysis will be performed when approximately 61% (730 events) of the positively adjudicated primary endpoint events have occurred. The final analysis will be performed at 100% (approximately 1200) of the positively adjudicated primary endpoint events, if the study is not stopped early. At the interim analysis time point (Figure GBDJ.4.1), superiority will be tested first; if successful, the trial may stop and superiority will be declared. Otherwise, the trial will continue to the end, where, at 1200 events, superiority will be tested followed by noninferiority. The interim and final analyses will be performed on unblinded study data by the Independent Data Monitoring Committee (IDMC) that will use the results and the foregoing decision rules as guidelines. Thus, if the interim analysis shows clear benefit of dulaglutide over placebo for the primary endpoint, the IDMC may recommend early termination of the study. Alternatively, if the boundaries are crossed at the interim analysis, the IDMC may still recommend the trial continue and not stop for early efficacy. At any time during the trial, the IDMC could recommend stopping the trial for safety reasons. The alpha used across the analyses will be monitored by an O'Brien-Fleming spending function (O'Brien and Fleming 1979; Jennison and Turnbull 2000), (for example, with 730 events at the interim, 2-sided alpha = 0.008; power 51.3%). The alpha used at the final analysis will be adjusted to maintain the overall type I error control at a 2-sided significance level of 0.05. This will be accomplished using EAST software to calculate the alpha level for the final analysis considering the actual amount of information at the interim analysis (for example, with 1200 events at the final analysis, 2-sided alpha = 0.048; overall power = 92.8%). At the final analysis, superiority will be tested followed by noninferiority. The adjusted 95% confidence interval (CI) for the hazard ratio will be calculated.

Interim Decision Rule
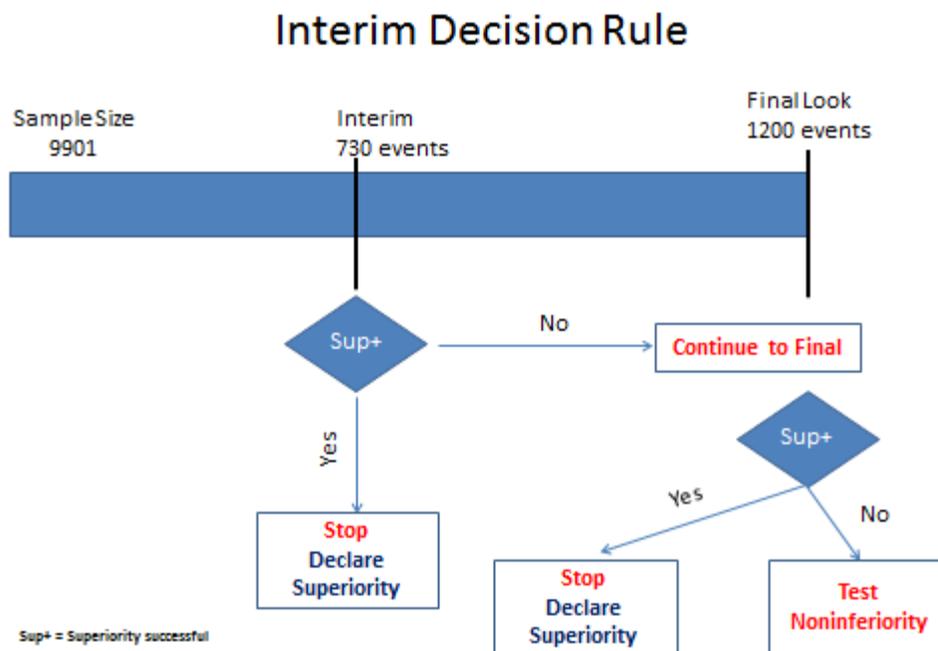
Figure GBDJ.4.1.   Interim decision rule.

The median unbiased estimator for the hazard ratio will be reported if the trial is stopped early for efficacy at an interim analysis.

4.1. Determination of Sample Size

For the initial powering of the trial, a minimum of 1067 unique primary endpoint events will be required to provide 90% power to demonstrate the superiority of dulaglutide over placebo at a true hazard ratio of 0.82 and 2-sided significance level of 0.05. To calculate the sample size, the following assumptions were used and yielded a sample size requirement of approximately 9600 patients: (1) two-sided significance level of 0.05; (2) 90% power for the primary endpoint; (3) patient accrual over 3 years; (4) annual placebo group event rate of 2.0% for the primary endpoint; (5) maximum duration of follow-up of 8 years; (6) a detectable hazard ratio of 0.82 between dulaglutide and placebo in terms of the primary endpoint; and (7) annual dropout rate of 0.15%.

The calculations were performed using nQuery Advisor® Version 7.0. This software provides sample size estimates for tests based on exponential survival, accrual period and dropouts. The sample size and other trial characteristics, such as interim analysis power, were also assessed through trial simulation. Trial assumptions were based on information from the scientific leadership of the study and a review of the relevant literature.

4.2. General Considerations

All entered data will be verified, and archived at a contract research organization (CRO) external to Lilly and/or at Lilly. An independent statistical analysis center (ISAC) will perform analyses for the IDMC prior to unblinding. After database lock at the conclusion of the study, analyses for the major key manuscripts will be conducted by the same or another ISAC based on data supplied by the CRO and the relevant manuscripts will be prepared by a writing group chosen by the Operations Committee. Data listings, summaries, and analyses will also be performed by the CRO and/or by Lilly for the purpose of the final clinical study report (CSR).

Efficacy and safety analyses will be conducted using the intent-to-treat (ITT) population. This population will include all randomized patients within the treatment group the patients were assigned to regardless of whether or not they took study drug or the correct study drug. A patient is considered randomized once the call has been made to interactive voice response system (IVRS) and a treatment is assigned at Visit 3.

Additional analyses will be conducted using the per-protocol (PP) population. The PP population is a subset of the ITT population defined as all randomized patients who have not permanently discontinued study drug, discontinued from the study, have an overall adherence of ≥75%, and have no important protocol deviations. The primary efficacy analysis will be repeated using the PP population. Important protocol deviations are defined in Section 4.7.

The analysis populations used in this study are defined in Table 4.1.

Unless otherwise specified, listings will be conducted using all randomized patients. The data collected will be presented as listings by investigator site, patient, and treatment.

Unless otherwise noted, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05 and CIs will be calculated at a 2-sided 95% confidence level. A graphical approach for multiple comparisons (Bretz et al. 2009; Bretz et al. 2011) will be used to strongly control the overall Type I error (2-sided alpha of 0.05) for testing the null hypothesis of no treatment effect with respect to the secondary endpoints.

For subgroup analyses, all tests of interactions between treatment groups and other factors will be conducted at a nominal 2-sided alpha level of 0.10 and will be deemed to be exploratory. No adjustment for multiplicity will be performed unless otherwise specified.

Countries in similar geographic regions with <10 patients will be pooled in order to achieve a pooled country of at least 10 patients. All analyses using country in the model will use pooled country, unless otherwise specified. The final pooling by country and geographic region will be specified prior to data lock.

The baseline is Visit 3. If baseline data are missing, the last measurement taken prior to this visit will be used for the baseline measurement.

The primary analysis and key secondary CV endpoint analyses will be based on adjudicated events that occurred after randomization (Visit 3). Sensitivity analyses may be conducted

including events reported by investigators or events identified through safety reviews, to assess their impact on the primary analysis results.

All analyses will be implemented using SAS Version 8.2® or higher.

4.3. Default Analysis Methods

A default analysis method will be defined for certain types of variables (for example, baseline continuous, baseline categorical and time-to-event) that will be commonly encountered in this study. If a specific analysis requires a methodology other than the default for that variable type, the methodology will be described in the appropriate sections of this SAP.

4.3.1. Baseline Analyses

Measurements collected at Visit 3 will be considered the baseline values. If data from Visit 3 are missing, measurements from the screening/run-in period (between Visit 1 and Visit 3) will be used. If a patient has no information for a variable prior to randomization, data will not be imputed and the patient will not be included in the analysis. Baseline analysis for any particular variable will be conducted using all randomized patients with baseline data for that variable.

The default analysis method for continuous baseline variables will be an ANOVA model with only a fixed effect for treatment; however, in situations where the baseline variable will be expected to violate the assumption of normality, Wilcoxon's rank sum test will be used. Such situations will be appropriately identified throughout this SAP. Summaries for continuous variables will include descriptive statistics (that is, number of patients, mean, standard deviation (SD), sample size, median, 10th percentile, 90th percentile, and interquartile range).

Categorical variables will be compared between treatments using Pearson's chi-square test if the expected count is at least 5 in at least 80% of the cells; otherwise, Fisher's exact test will be used. Summary statistics for categorical variables will include sample size, number, and proportion of patients.

For the CSR, a baseline factor will be considered as a covariate for a covariate- adjusted analysis only if the baseline factor is both clinically and statistically significantly different between treatment groups.

4.3.2. Post-Baseline Analyses of Continuous Variables

For each continuous response variable, visit-specific analyses will be performed for the visits where the variable was scheduled to be measured. If a patient has no post-baseline measurements of a variable, the patient will not be included in the visit-specific analyses of that variable. Analyses of change and percent change from baseline to specific visits will have the same requirements plus the additional constraint that the patient has a baseline measurement of the variable.

The default analysis model for continuous response variables expected to be normally distributed will be a mixed-effects model for repeated measures (MMRM) using restricted maximum

likelihood (REML). The MMRM model will be used to analyze changes from baseline with the baseline value as the covariate. The model will include fixed effects for treatment, visit, treatment-by-visit interaction, the baseline as a covariate, and the patient as a random effect. An unstructured covariance structure will be used to model the within-patient errors. This analysis includes visits where the continuous response variable was scheduled to be measured.

If this analysis fails to converge, the following covariance structures will be tested in order:

- toeplitz with heterogeneity

- autoregressive with heterogeneity, by visit

- compound symmetry with heterogeneous variances, by visit

- toeplitz,

- autoregressive

- compound symmetry without heterogeneous variances, by visit

The first covariance structure that converges will be used. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. Significance tests will be based on least-squares mean (LS Mean) using Type III sum of squares.

Summary statistics will include sample size, mean, SD, median, 10th and 90th percentiles for both the actual and the change from baseline measurements. Least-squares mean (LS Mean) and standard error (SE) derived from the model will also be displayed for the change from baseline measurement. Treatment comparisons will be displayed showing the treatment difference LS Mean and the 95% confidence limits along with the p-value.

For continuous lab measurements, an analysis of variance (ANOVA) on rank-transformed data will be used and p-values for the difference between dulaglutide and placebo will be reported, unless stated otherwise. The model includes treatment. Treatment group comparisons will be performed with no multiplicity adjustment. Categorical laboratory measures will be analyzed using Chi-square test or Fisher's exact test.

In situations where the continuous response variable will be expected to violate the assumption of normality, Wilcoxon's rank sum test will be used. Such situations will be appropriately identified throughout this SAP. Between-treatment comparisons will be assessed using the p-value, and within-treatment changes will be assessed using the p-value from the signed rank test. Summaries for these variables will include descriptive statistics: number of patients, median, and 25th and 75th percentiles. In addition, the Hodges-Lehman estimator of the treatment contrast will be provided.

4.3.3. Post-baseline Analyses of Categorical Variables

Categorical variables will be compared between treatment groups using the Pearson chi-square test if the expected count is at least 5 in at least 80% of the cells; otherwise, Fisher's exact test will be used. Summaries for these variables will include the between-treatment p-value and the

descriptive statistics: number and proportion of patients. For post-baseline outcome variables, the relative risk estimate and the associated 95% CI will be provided, if there are at least 10 patients with the outcome.

If the categorical variable is associated with a scheduled post-baseline measurement, then a patient must have a measurement of the variable to be included. However, when the categorical variable is the occurrence of an event (yes/no), such as development of an adverse event (AE), then all randomized patients will be included in the analysis.

4.3.4. Time-to-Event Analyses

Time-to-event analyses will be performed for each of the adjudicated outcomes. For each analysis, all adjudicated events in the locked database will be used.

Time-to-event variables will be analyzed using survival analysis methodology if the total number of outcomes is 10 or more. A Cox proportional hazards regression analysis, where the model only includes a fixed effect for treatment, will be used to derive the hazard ratio (dulaglutide/placebo) and the associated 95% CI. The between-treatment comparison will be based on the p-value from the Cox model. The proportional hazard assumption will be examined graphically. If not met, data will be analyzed using accelerated failure time models. Kaplan-Meier (KM) estimates of the survival curve for each treatment will be generated. The number and proportion of patients with the event will be provided, along with the between-treatment p-value. Tied event times will be handled by the Exact method.

If the number of outcomes is <10, survival analyses will not be performed. Instead, Fisher's exact test will be used and the summary statistics will include the number and proportion of patients with the event plus the between-treatment p-value.

For adjudicated outcomes, the incidence rate per 100 person-years of follow-up will be calculated for each treatment group. The numerator will be the number of patients with the event, and the denominator will be the event-specific total person-years of follow-up divided by 100. Total person-years of follow-up is the sum, over patients, of the time on study until the first outcome (first event time or censoring time). The absolute risk difference (ARD) will then be calculated by subtracting the incidence in the dulaglutide arm from that in the placebo arm. For some analyses, the number needed to treat (NNT) to prevent an additional event will be derived using the incidence rates, that is, 1/(Placebo rate – dulaglutide rate). These analyses will be performed where documented in the SAP, and this statistic will only be derived if the p-value from the Cox model is statistically significant.

4.3.4.1. Time to Event

For each patient, time-to-event for an event of interest will be the number of days between the date of randomization and the onset date of the event plus 1 day if the patient experiences the event or the number of days between the date of randomization and the censoring date (Section 4.3.4.3) plus 1 day if the patient does not experience the event. If a patient experiences

multiple events (for example, multiple strokes) the date of the first event will be used, unless otherwise specified.

4.3.4.2. Person-Years of Follow-up

Person-years of follow up for an event of interest will be calculated for each patient as the time to event (defined in Section 4.3.4.1) divided by 365.25.

The total person-years of study follow-up will be calculated for each patient as the number of days between the randomization date and the censoring date plus 1 day divided by 365.25. The censoring date is defined in Section 4.3.4.3 for time-to-event analyses (other than mortality analysis).

4.3.4.3. Censoring Date

For time-to-event analyses (except for mortality analyses), the censoring date for a patient is the Final Visit date if a Final Visit was conducted, the discontinuation date if the patient discontinues from the study early, or the patient's date of death if the patient dies during the course of the study. This censoring date will be used in all analyses (except the mortality analyses) to censor patients who have not experienced the event of interest.

For time-to-event analyses for mortality, the censoring date for a patient is the Final Visit date if the patient is known to be alive at the time of the Final Visit. If the patient discontinues from the study early, the censoring date will be the last date that the investigator can ascertain the patient was alive.

4.3.4.4. Handling of Missing Dates

For all adjudicated events, the date the Clinical Endpoints Committee (CEC) indicates the event occurred will be used in all analyses. If this date is missing then the investigator-reported date will be used.

For an incomplete endpoint event (that is, primary endpoint or secondary efficacy endpoint events) date, imputation will be performed as outlined below:

- If only the day of the event date is missing, the day will be imputed as the 15th of the reported month.

- If both the month and day of the event date are missing, the month and day will be imputed as 30 June of the reported year.

- If only the month is missing, the month will be imputed as June.

- In the case that the imputed event date falls after the patient's censoring date as defined above, the incomplete event date will be imputed as the censoring date (for example, if an incomplete onset date of an event is X-X-2012 and the patient's censoring date is 05-22-2012, then the onset date of the event will be imputed as 05-22-2012 rather than as the date of 06-30-2012 as imputed by following the procedure stated above for missing month and day.

For an incomplete death date, imputation will be performed as described below:

- If only the day of the death date is missing, the day will be imputed as follows.

    - (a) If the date of the last reported contact for the patient falls in the same month and year as the death date where the day is missing, the day will be imputed to fall halfway between the last reported contact and the end of the given month (for example, if an incomplete death date is 04-X-2012, and the date of the last reported contact is 04-22-2012, the death date will be imputed as 04-26-2012);

    - (b) If the date of the last reported contact for the patient occurs before the reported month and year of the death date, the day will be imputed as the 15th of the reported month (for example, if an incomplete death date is 04-X-2012, and the date of the last reported contact is 03-26-2012, the death date will be imputed as 04-15-2012).

- If both the month and day of the death date are missing, the month and day will be imputed as follows.

    - (a) If the date of the last reported contact for the patient falls in the same year as the incomplete death date, the death date will be imputed as the first of the month falling halfway between the month of the last reported contact and the end of the year (for example, if an incomplete death date is X-X-2012, and the date of the last reported contact for the patient is 06-22-2012, the death date will be imputed as 09-01-2012).

    - (b) If the year of the last reported contact date for the patient occurs before the year of the incomplete death date, the death date will be imputed as 30 June of the reported year (for example, if an incomplete death date is X-X-2012, and the date of the last reported contact is 06-22-2011, the death date will be imputed as 06-30-2012).

4.3.5. Subgroup and Risk-Adjusted Analyses

Subgroup analyses and analyses that account for differences in baseline risk factors will be performed and will be regarded as exploratory. The subgroup analyses for time-to-event variables will only be conducted if the number of outcomes is at least 50.

4.3.5.1. Prespecified Subgroup Analyses

Subgroup analyses will be performed for the following prespecified subgroup variables:

- gender (female and male)

- age group (age <median, and age ≥median)

- duration of diabetes (duration <5 years, 5 years ≤ duration <10 years and duration ≥10 years)

- body mass index (BMI) (BMI <median and BMI ≥median)

- baseline HbA1C (HbA1C <median and HbA1C ≥median)

- geography (region) (US/Canada, South America, Europe, and Asia Pacific)

- prior CV event (Yes, No)

For the purpose of the subgroup analyses, South Africa will be included in the geographical subgroup of Europe, Australia and New Zealand in the subgroup of Asia Pacific, and Mexico in the subgroup of South America.

A prior CV event is defined as a history of MI, or a history of myocardial ischemia by a stress test or with cardiac imaging, ischemic stroke, coronary, carotid or peripheral artery revascularization, or hospitalization for unstable angina with electrocardiogram (ECG) changes, or need for percutaneous coronary intervention (PCI).

4.3.5.2. Exploratory Subgroup Analyses

To select additional factors for subgroup analyses, clinically relevant baseline characteristics will be identified for the primary outcome variable and each of its components. A Cox proportional hazards regression model adjusted for baseline risk factors will be obtained using a model selection process, guided by clinical review of the published literature. All these factors will be forced into the model, along with the above prespecified subgroup variables. For all factors, the significance level for remaining in the model will be 0.05. After the selection of terms into the model concludes, treatment will be added to the model. For all dichotomous terms in the final model, parameter estimates, standard errors and p-values plus the hazard ratio and its 95% CI will be reported.

Any factors remaining in the final proportional hazards model will also be used for subgroup analyses.

4.3.6. *Multiple Comparisons/Multiplicity Adjustments*

The primary objective of this study is to demonstrate with statistical significance a clinically meaningful reduction in the incidence of primary endpoint events. There will be 1 interim analysis and 1 final analysis. The interim analysis will be performed when approximately 61% (730) of the primary endpoint events have occurred and have been adjudicated as such. The final analysis will be performed at 100% of the total information (approximately 1200 adjudicated primary endpoint events) if the study is not stopped early. Appropriate adjustments for multiplicity will be made to maintain an overall type I error rate of <0.05 (Section 4).

Multiplicity adjustments will be performed for the analyses on the primary and secondary efficacy endpoints in order to control the overall Type I error rate at a 2-sided alpha level of 0.05. For repeated testing of the primary outcome via interim analysis, an O'Brien-Fleming alpha spending function will be used to control alpha across the interim and final analyses for the testing of the primary endpoint (GBDJ Protocol Section 12.2.2). A graphical approach will be

used to control the type I error rate across the secondary endpoint analyses at the final analysis time point or at the interim analysis time-point if the decision is to stop the trial following the interim analysis (GBDJ Protocol Section 12.2.2 and Section 12.2.12 for the Interim Decision Rule).

4.3.6.1. Graphical Approach

A graphical approach using a sequentially rejective Bonferonni multiple-testing procedure to control the family wise error rate (FWER) will be used for the secondary hypotheses in this study. In general, a graphical approach is characterized by prespecifying:

- the hypotheses (or nodes) within the testing algorithm

- an initial allocation of α

- weights along the lines (edges) that connect the hypotheses tests.

The initial allocation of α is such that the sum of all α is the full alpha available for the analyses. The algorithm begins by testing all hypotheses that have α allocated to them. If any hypothesis is rejected then the graph is updated to reflect the reallocation of α where the proportion of α reallocated to other hypotheses is determined by the weights along the edges. Hence the α can be considered "recycled" as described by Bretz and colleagues (2009) as long as at least one hypothesis is rejected. This iterative process of updating the graph and reallocating the α is repeated until no further hypotheses can be tested.

Figure GBDJ.4.2 shows the primary and secondary endpoints, the α allocation, and associated weights for the REWIND Study. Only the hypothesis test for the primary endpoint (Composite MACE) is initially allocated the full value of α at the given analysis time-point (for example, at the final analysis time-point with 1200 events, $\alpha = 0.048$, 2-sided); all other hypotheses are allocated $\alpha=0$. Additionally the weights are provided along the edges.

The first hypothesis tested will be for the primary endpoint (time to first composite MACE) which is tested at the alpha value used for the primary analysis at the given analysis time point. If that hypothesis is rejected then the graph is updated and the hypotheses for the following 3 endpoints will be tested with an α based on the weights indicated by the 3 edges (arrows) coming from the Composite MACE. Figure GBDJ.4.3 demonstrates what the graph would look like if the first test (Composite MACE primary endpoint) is rejected at the final analysis time-point with 1200 events, with α reallocated as follows:

- Time to first Nonfatal MI has a weight of 0.15 resulting in $\alpha = (0.048*0.15) = 0.0072$

- Time to CV Death has a weight of 0.7 resulting in $\alpha = (0.048*0.7) = 0.0336$

- Time to first Nonfatal Stroke has a weight of 0.15 resulting in $\alpha = (0.048*0.15) = 0.0072$.

If any of the 3 hypotheses above are rejected, the graph is again updated and its alpha is reallocated to the subsequent tests according to the weights given in the figure. It should be noted it is possible for the hypothesis for a particular endpoint to fail to be rejected the first time it is tested, but subsequently receive enough α from successfully rejected hypotheses and be rejected the second time it is tested. This iterative process of updating the graph and redistribution of α is continued until no further hypotheses can be tested.

The graphical approach to multiplicity control has been evaluated in several peer reviewed publications: Bretz and colleagues (2009), Bretz and colleagues (2011); and Alosh and colleagues (2014). Because graphical approaches are closed testing procedures, they control the FWER strongly (Aloshet al. 2014). This closed testing system controls the FWER to ≤ 0.05.
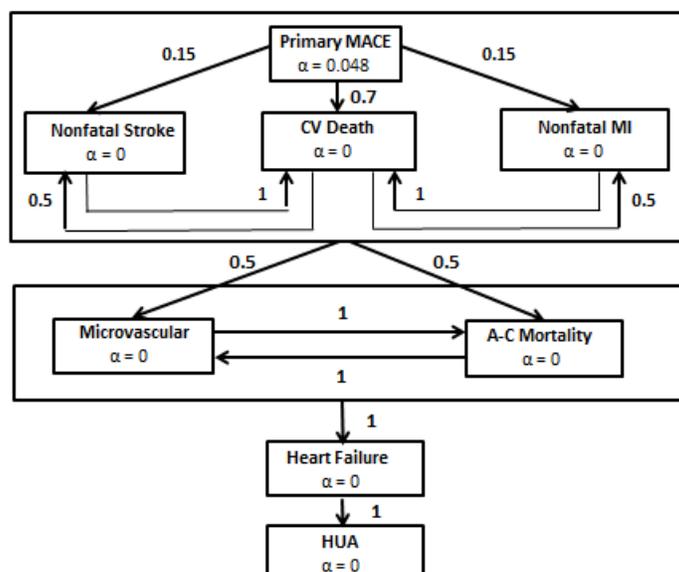
Abbreviations: A-C Mortality = all-cause mortality; CV = cardiovascular; HUA = hospitalization for unstable angina; MACE = major adverse cardiovascular event; MI = myocardial infarction.

Figure GBDJ.4.2.   **The graphical approach with initial α allocation and weights at the final analysis time point with 1200 primary MACE events.**
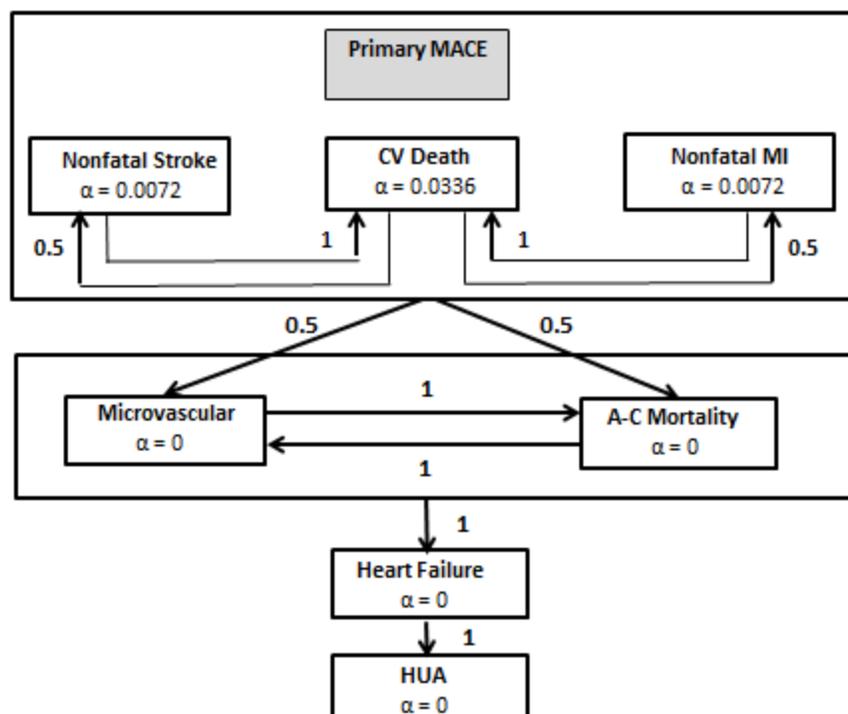
Abbreviations: A-C Mortality = all-cause mortality; CV = cardiovascular; HUA = hospitalization for unstable angina; MACE = major adverse cardiovascular event; MI = myocardial infarction.

Figure GBDJ.4.3. **The graphical approach reflecting α propagation after rejecting the first hypothesis at the final analysis time point with 1200 primary MACE events.**

No adjustments will be made for multiple comparisons on other endpoints.

4.3.6.2. Interim Analysis and Control of Type I Error following an Efficacy Stop Decision

4.3.6.2.1. *Data Cut-off Date for the Interim Analysis*

The interim analysis is planned to be conducted when a minimum of 730 positively adjudicated MACE endpoints events have occurred. It is expected that there will be more than 730 events in the pool for the interim analysis, as all events that are adjudicated on the same day as the 730th positive event will also be included in the analysis. The data cut-off date for the interim analysis will be the event date, from the adjudicator's opinion, of the most recent event in this pool of unique subjects with positively adjudicated primary MACE endpoint events. In addition, all efforts will be made to adjudicate the majority of all MACE events with events dates prior to this cut-off date, and these adjudicated events will be included in the interim analysis (Figure GBDJ.4.4).

At the interim analysis, the alpha level used for the analysis will be determined based on the actual number of events included in the analysis using EAST Software (EAST 6, Cytel, www.cytel.com). If the trial continues to the end following an IDMC review of the interim

results, the final alpha will be adjusted based on the final number of primary MACE events using EAST software. If the IDMC recommends stopping the trial for efficacy following the interim analysis, analyses will be performed on this post-interim, final-lock database using the adjusted significance level alpha based on the Whitehead procedure (Section 4.3.6.2.2).

Example for Interim Database Cut-off

For the primary analysis, include positively adjudicated events that occurred by the cut-off date

Adjudication Date

30Nov2011 — 10Sept2011 Event 1 | 22Sept2011 Not Event | 1Oct2011 Event 2 ← Date Event Occurred (as stated by adjudicators)

21Dec2011 — 21Nov2011 Not Event | 16Sept2011 Event 3 | 2Oct2011 Event 4 | 14Oct2011 Not Event

Cut-off Date: 18Jul2016

Step 2: 18Jul2016 is the date the latest event occurred which was adjudicated by 22Aug2016

15Aug2016 — 24Apr2016 Event 727 | 8May2016 Event 728 | 13Apr2016 Not Event | 18Jul2016 Event 729

22Aug2016 — 28May2016 Not Event | 12May2016 Event 730 | 6Jun2016 Event 731

Step 1: 22Aug2016 is the adjudication date for the 730th event

Note: Dates in the figure are for illustrative purposes only.

Figure GBDJ.4.4. **Example of data cut-off date for the analysis of the primary MACE events.**

4.3.6.2.2.   Controlling Type I Error across Interim and Final Analyses Following Early Termination for Efficacy

All events until study closure will be collected and analyzed as part of the conduct of the study following a decision to stop the study early for efficacy. This post-interim, final-lock dataset will be comprised of all events, including those events collected after the interim stopping decision until final trial closure. These data are valuable in the full interpretation of the conduct of the study and will be included in the submission to Food and Drug Administration (FDA). If the study is stopped at the interim as a result of meeting the interim boundary, the interim analysis data used for making the decision will be included as part of the Common Technical Document (CTD). The post-interim final lock dataset will be the primary database to be used to support safety and efficacy in the CTD and for proposed labeling. While the Type I error rate at interim is already controlled at the interim lock, in order to control for Type I error for the post-interim final analysis, Whitehead's approach will be used to recalculate the rejection boundaries since the additional events after the interim will be included in the primary dataset (Whitehead1992). Both the interim and post-interim final data (primary and key secondary endpoints) will be included in the CTD and the CSR for efficacy.

The following is an example of the use of Whitehead's approach in the context described above. For example, with 730 events at the interim lock, the one-sided p-value boundary is 0.004. If the interim p-value is less than the stopping boundary, then the trial may be stopped early for

efficacy. If the trial is stopped early for efficacy, additional events will accrue between the interim lock and the post-interim final lock and this could range between 100 to 250 additional events. Whitehead's approach controls Type I error by recomputing the final p-value boundary based on the number of events included in the interim lock plus the additional events accrued. For example, with an additional 250 events following the interim lock (for a total of 980 events), the one-sided p-value boundary for the primary analysis is 0.013. Using Whitehead's approach, at the post-interim final analysis (following trial stop), the stopping boundary will be calculated based on the total number of primary MACE events included in the post-interim final lock (out of the 1200 events planned for the end of the trial) using the protocol-specified O'Brien-Fleming spending function. If the trial does not stop early for efficacy, Whitehead's approach will not be used.

4.4. Multicenter Study

Patients will be enrolled at approximately 486 investigational sites in approximately 27 different countries. To control for differences between sites and assure that treatment allocation is balanced within site, randomization will be stratified by site. Randomization blocks of size 4 will be dynamically allocated to sites. Patients will be assigned in a 1:1 ratio to the next available treatment (either dulaglutide 1.5 mg/week or placebo) from the block currently allocated to their site at the time the call was made to the IVRS. The number of patients randomized at each site in each country will be provided.

4.5. Patient Populations/Analysis Subsets

The patient populations and analysis subsets used in the study are described below.

Table 4.1. **Analysis Population for REWIND**

Population	Definition
All Entered	All patients who signed informed consents
All Randomized	All patients who were randomized to a treatment arm
Non-Randomized	All patients entered but not randomized to a treatment arm
Intent-to-Treat	All patients randomized within their treatment group regardless of whether or not they took study drug or correct study drug (same as all randomized population)
Per-Protocol	All patients in ITT and also meet the following criteria • have not permanently discontinued study drug • no important protocol deviations • have completed the study • have an overall adherence with study drug of ≥75%

4.6. Patient Disposition

A listing of patient discontinuations will be presented for all randomized patients. Summary analyses will be conducted for the ITT and PP populations.

Number and proportions of patients will be presented for each treatment group and compared across treatment groups using a Chi-square test or Fisher's exact test.

4.7. Important Protocol Deviations

Important protocol deviations will be listed for all randomized patients. The following protocol deviations will be considered important:

- not meeting any of the inclusion criteria [1], [6], and [9], or meeting any of the exclusion criteria [10], [11], [12], [13], and [21] (REWIND study protocol) that affect the primary endpoint analysis

- nonadherence to treatment regimen (that is, overall treatment adherence <75%)

- incorrect dispensation of study drug (patient received incorrect treatment)

- unblinding of treatment assignment for any reason

- using prohibited concomitant medications (that is, non-study GLP-1 analogs, pramlintide, or prescription weight loss drugs)

- safety laboratory or measures (calcitonin and/or serum creatinine and/or ECG) not taken for 18 or more months and study drug not discontinued

4.8. Patient Characteristics at Baseline

Demographic and baseline characteristics will be summarized by treatment group using ITT and PP populations. For continuous measures, summary statistics will include sample size, mean, median, 10th percentile, 90th percentile, and standard deviations. Mean will be analyzed using ANOVA. For categorical measures, summary statistics will include sample size, number, and proportion. Treatment group comparisons will be performed using a Chi-square test or Fisher's exact test.

4.9. Concomitant Medications

Concomitant medications will be summarized by different categories and treatment group using ITT population. All concomitant therapies originally mapped using the WHODRUG dictionary in the clinical trial database will be further classified using Anatomical Therapeutic Chemical (ATC) code for reporting purpose. The number and proportion of patients will be analyzed using a Chi-square test or Fisher's exact test.

4.10. Historical Illnesses at Visit 1

A limited number of historical illnesses collected at Visit 1 will be listed using all randomized patients. Summary reports will be conducted by treatment group using the ITT population. Historical illnesses will be reported using the terms in the medical history questions on the case report forms. The number and proportion of patients will be analyzed using a Chi-square test or Fisher's exact test.

4.11. Treatment Adherence

Treatment adherence will be assessed for each visit interval. Patients will be instructed to return any unused study drug syringes at each study visit for the purposes of study drug accountability. Study drug adherence will be calculated at each visit after baseline when study drug is dispensed based on the percentage of syringes used, specifically it will be calculated as follows:

Study drug adherence for each visit =
        [(no. of syringes dispensed – no. of syringes returned) /
        (no. of weeks between the 2 consecutive visits)]   *   100%.

A patient will be considered adherent for each visit interval if he/she uses at least 75% of the study drug syringes dispensed for that interval.

Treatment adherence will be listed and summarized using the ITT population. The number and proportion of patients who are compliant at each visit by treatment group will be summarized and compared using a Chi-square test unless the total count is <10.

The overall adherence during the study will be calculated for each patient. This will be calculated by taking the number of visits the patient was compliant divided by the total number of visits with nonmissing adherence data for this patient (that is, the proportion of visits at which the patient was compliant among visits with non-missing adherence data for the patient). The overall adherence will be summarized and presented in descriptive statistics that include the sample size, mean, median, 10th and 90th percentiles, and standard deviation. The overall adherence will be used as one of the factors when determining if a patient is eligible for the PP population.

4.12. Treatment Exposure and Study Duration

Patients may discontinue study treatment but remain in the study.

Treatment exposure will be defined as the total duration of time that the patient is believed to be taking study drug. This time period will be derived by subtracting the randomization date (Visit 3) minus 1 day from the last visit date for patients actively participating in the study at the time of database lock (at the conclusion of REWIND) or the discontinuation date for patients who discontinue study drug. The duration of treatment exposure will be reported in months and will be compared between treatments using nonparametric methods (Wilcoxon rank-sum tests). The duration of treatment exposure will be summarized by treatment group using the ITT population. The duration of exposure will be categorized into the following groups: ≥2 weeks, ≥3 months (13 weeks), ≥6 months (26 weeks), ≥1 year (52 weeks), ≥2 years, ≥3 years, ≥4 years and ≥5 years, etc. These categories will be summarized by the number and proportion of patients in each category by treatment group.

Keeping with the ITT principle, the study duration for a patient will commence on the date of randomization to the time of last participation in the study. This will be the patient's safety reporting period. At the time of datalock (at the conclusion of REWIND), the time of last participation will be the Final Visit date for patients still actively participating in the study; the

date of discontinuation if the patient discontinued from the study early; or the date of death if the patient dies during the study. The study duration will be reported in months and will be compared between treatments using nonparametric methods (Wilcoxon rank-sum tests). In addition, a categorical breakdown of the study duration will be provided using the same categories as the duration of treatment exposure.

The number and proportion of patients who discontinue study drug will be compared between treatment groups with separate analyses for the reasons for the discontinuation, (for example, AE, protocol, or patient decision). The length of time patients are off study drug will be summarized by treatment group. The analyses for discontinuations will also be performed by visit.

4.13. Primary Efficacy Measure

The primary efficacy measure is the time to first occurrence (after randomization) of a composite of death from CV causes, nonfatal MI, or nonfatal stroke.

An independent Clinical Endpoints Committee (CEC) will adjudicate all primary endpoint events. The CEC Charter will contain the final detailed event definitions used for adjudication; however high level definitions of each primary endpoint event are provided below.

1) **Death from Cardiovascular Causes** will be defined as a death resulting from an acute MI, sudden cardiac death, death due to heart failure, death due to stroke, and death due to other CV causes. All cases in which the cause of death cannot be determined (that is, undetermined) will be included in deaths from CV causes.

2) **Myocardial Infarction (MI).** The term MI will be used when there is evidence of myocardial necrosis (that is, changes in cardiac biomarkers or post mortem pathological findings) in a clinical setting consistent with myocardial ischemia. Myocardial infarction will include the following subtypes: spontaneous MI, PCI-related MI, coronary artery bypass grafting (CABG)-related MI, and silent MI.

3) **Stroke** will be defined as an acute episode of neurological dysfunction caused by a focal or global brain, spinal cord, or retinal vascular injury. Strokes will be classified as ischemic, hemorrhagic, or undetermined. Stroke disability, as measured using the modified Rankin scale, will be assessed at approximately 30 days after the diagnosis.

The analysis for the primary efficacy measure will be based on the ITT population.

4.13.1. Primary Analysis Model

The primary analysis model is a Cox proportional hazards regression model. The model includes treatment as a fixed effect.

4.13.2. Analysis of Primary Endpoint

The endpoint for the primary analysis is defined as the first occurrence after randomization of death due to a CV cause, nonfatal MI, or nonfatal stroke (adjudicated as such). Time-to-event analyses (Section 4.3.4) will be performed for the composite primary endpoint. Treatment comparisons will be based on the hazard ratio and its 95% CI from the Cox model.

Counts and proportions of patients who experience a primary endpoint event will be calculated. Person-years of follow-up for the primary endpoint, and the incidence rate, calculated by dividing the number of patients who developed the event during the study period by the event-specific person-years of follow-up, will be provided. The ARD for an endpoint will be calculated based on the difference in cumulative incidence between the 2 treatment groups at the end of the study period. The ARD and NNT (Section 4.3.4) will be calculated for the primary endpoint. An adjusted 95% CI for the hazard ratio (dulaglutide/placebo) from the Cox proportional hazards regression model will be provided.

The primary analysis at the conclusion of the trial will be a superiority comparison of dulaglutide versus placebo. If the superiority test fails, then a noninferiority test with a 1.3 margin will be performed. If the upper limit of the 95% CI is below 1.0 (after adjustment for interim looks), dulaglutide will be declared superior to placebo in reducing the incidence of CV events. If the upper limit of the adjusted 95% CI of dulaglutide versus placebo is above 1.0 but below 1.3, dulaglutide will be declared noninferior to placebo in its effects on CV events.

4.13.3. Subgroup Analysis of Primary Endpoint

The effects of dulaglutide compared to placebo on the incidence of the composite primary endpoint events will be examined across the prespecified subgroups of interest listed in Section 4.3.5.

The clinically relevant baseline characteristics for exploratory subgroup analyses for the primary endpoint include: prior MI, prior history of stroke or transient ischemic attack (TIA), prior hospitalization for unstable angina, prior coronary revascularization, existing microalbuminuria (defined as a urine albumin/creatinine ratio (ACR) >30 mg/g and ≤300 mg/g) or macroalbuminuria (see Section 4.17.3), current smoker, hypertension, hyperlipidemia, aspirin use, beta-blocker use, calcium channel blocker use, ACE inhibitor or angiotensin receptor blocker use, diuretic use, and antithrombotic (nonaspirin antiplatelets, vitamin K antagonists, heparin, direct thrombin inhibitor, or other antithrombotics).

A risk-adjusted analysis will be performed using these baseline characteristics, the prespecified subgroups, and a multifactor proportional hazards model. The baseline characteristics remaining in the final multifactor proportional hazards model plus the prespecified subgroup variables will be used for subgroup analyses.

The risk-adjusted and subgroup analyses will be performed using the ITT population.

LY2189265

4.13.4. Sensitivity Analyses of Primary Endpoint

The time-to-event analysis for the primary endpoint will be repeated using the PP population, but NNT will not be calculated.

Randomization was stratified by site. To account for the stratification, a time-to-event analysis of the primary endpoint stratified by site will be performed as a sensitivity analysis.

4.13.5. Assessing the Impact of Missing Data on the Primary Endpoint Analysis

Vital status or status on the primary endpoint may not be available for patients who discontinued early from the study or patients who were lost to follow-up during the course of the study. Sensitivity analyses will be performed exploring the impact of missing data on the primary endpoint findings. Specifically, 2 analyses of the primary endpoint will be performed. The first analysis which is extreme, will assume that all such patients in the dulaglutide group have experienced MACE primary endpoint events at the time of their discontinuations from the study or their last contact dates if they were lost to follow-up, while such placebo patients did not experience an event and thus are censored at their corresponding times. The second analysis, less extreme, will assume that all these patients from both treatment groups have experienced primary MACE endpoint events at the time of their discontinuations from the study or their last contact dates if they were lost to follow-up.

4.14. Secondary Efficacy Measures

Secondary efficacy measures include time (after randomization) to:

- first occurrence of the composite microvascular endpoint of diabetic retinopathy

- requiring laser therapy, vitrectomy, or anti-VEGF therapy; development of clinical proteinuria, a 30% decline in eGFR, or need for chronic renal replacement therapy

- first hospitalization for unstable angina

- first occurrence of each component of the composite primary endpoint

- death

- first occurrence of HF requiring hospitalization or an urgent HF visit

The independent CEC will adjudicate all deaths and hospitalizations for HF or unstable angina. The CEC Charter will contain the final detailed event definitions used for adjudication; however, high level definitions for these endpoints are provided below.

1) **All Cause Mortality** will be defined as deaths from CV causes, deaths from non-CV causes (for example, pulmonary, renal, etc.) and deaths not attributable to a CV or non-CV cause (that is, undetermined).

2) **Heart failure requiring hospitalization** will be defined as new or worsening clinical symptoms and physical signs of HF that require hospitalization for additional/increased therapy. An **urgent HF visit** will be defined as an urgent, unscheduled office/practice or emergency department visit (requires clinical signs and symptoms of HF and need for additional/increased therapy).

3) **Hospitalization for unstable angina** will be defined as clinical symptoms of myocardial ischemia (new or worsening) that necessitates hospitalization and one of the following: new or worsening ST or T wave changes on electrocardiogram (ECG), evidence of myocardial ischemia on imaging, angiographic evidence of a lesion in a coronary artery responsible for symptoms, need for coronary revascularization procedure (PCI or CABG) during the hospitalization, AND no evidence of an acute MI.

For the composite microvascular endpoint, the following definitions will apply:

1) **Diabetic retinopathy requiring laser therapy** will be defined as use of laser therapy (photocoagulation) for the treatment of diabetic retinopathy.

2) **Vitrectomy** for the treatment of diabetic retinopathy will be defined as a surgical procedure to remove the vitreous gel from the inside of the eye, and silicone gas, oil or other fluid is injected to fill the space the vitreous once occupied.

3) **Anti-VEGF therapy** for the treatment of diabetic retinopathy will be defined as an intravitreal injection(s) of an anti-VEGF agent for the treatment of diabetic retinopathy.

4) **Clinical proteinuria (macroalbuminuria)** will be defined as an ACR >300 mg/g (>33.9 mg/mmol).

5) **Renal replacement therapy (RRT)** will be defined as chronic hemodialysis or

6) Peritoneal dialysis used as maintenance therapy in patients with end stage renal disease (ESRD), or renal transplantation.

7) **A sustained 30% decline in eGFR** will be based on a 30% reduction from the

8) Baseline value (Visit 3) in 2 consecutive calculations of postrandomization eGFR, using the modification of diet in renal disease (MDRD) equation.

Events of laser therapy, vitrectomy, anti-VEGF therapy, or RRT will be prospectively collected. Identification of clinical proteinuria will be based on reported laboratory data (and/or calculated if needed) and eGFR will be calculated using reported laboratory (serum creatinine) and clinical data.

An event of the composite microvascular endpoint is the first occurrence after randomization of an event of diabetic retinopathy requiring laser therapy, vitrectomy for the treatment of diabetic retinopathy, anti-VEGF therapy for the treatment of diabetic retinopathy, development of clinical

proteinuria (macroalbuminuria, see Section 4.17.3), a 30% decline in eGFR (see Section 4.17.3) from baseline, or need for renal replacement therapy.

Time-to-event analyses will be performed for each of the secondary efficacy measures including the individual components of the composite primary endpoint using the ITT population. The analyses of death from CV causes, nonfatal MI, nonfatal stroke, HF requiring hospitalization or hospitalization for unstable angina will be based on adjudicated events. The NNT statistic will be provided for each analysis provided that the p-value from the Cox model is statistically significant. The incidence rate per 100 person-years of follow-up will also be calculated for each type of event.

4.14.1. Additional Analyses of All Cause Mortality

Deaths will be analyzed on the basis of the adjudicated cause of death which will be categorized into CV death and non-CV death. Deaths with an undetermined cause will be included in CV deaths for analysis purposes. Time-to-event analyses will be performed for all deaths as well as for each adjudicated cause of death (CV death versus non-CV death and the subcategories under CV death). Incidence rates per 100 person-years of follow-up will be calculated for each type of death for each treatment group. The NNT statistic will not be provided for these analyses. These analyses will be performed using the ITT population.

A by-patient listing of all deaths and the adjudicated outcome will be provided.

4.14.2. Additional Analyses of MI Endpoint

The additional analyses of the MI endpoint will be performed using the ITT population.

4.14.2.1. Analyses of Classifications MI Endpoint

4.14.2.1.1. Universal MI Definition

For each MI event, the adjudicators will classify the MI type using the universal MI definition (Thygesen et al. 2007) as follows:

Type 1: Spontaneous MI related to ischemia due to a primary coronary event such as plaque erosion and/or rupture, fissuring, or dissection

Type 2: MI secondary to ischemia due to either increased oxygen demand or decreased supply, for example, coronary artery spasm, coronary embolism, anemia, arrhythmias, hypertension, or hypotension

Type 3: Sudden unexpected cardiac death, including cardiac arrest, often with symptoms suggestive of myocardial ischemia, accompanied by presumably new ST elevation, or new Left Bundle Branch Block, or evidence of fresh thrombus in a coronary artery by angiography and/or at autopsy, but death occurring before blood samples could be obtained, or at a time before the appearance of cardiac biomarkers in the blood

Type 4: MI associated with PCI or MI associated with stent thrombosis as documented by angiography or at autopsy

Type 5: MI associated with CABG

Each type of MI (using the universal MI definition) will be further summarized in categories of multiples (1-2X, 2-3X, 3-5X, 5-10X, and >10X) of the 99th percentile of the upper reference limit (URL) of cardiac biomarkers creatine kinase-MB (CK-MB) or troponin, and compared between treatment groups.

4.14.2.1.2. Other Classifications of Myocardial Infarction

The adjudicators will also classify MI type as STEMI vs NSTEMI and by subtypes of MI (Spontaneous, Periprocedural [PCI or CABG related], or Silent). Myocardial infarction types will be summarized using number and proportion of patients for both treatment groups using these different classification schemes. For these analyses, the denominator will be the total number of patients having an MI. If a patient has >1 MI, only the first MI will be included in the analysis.

4.14.2.2. Subgroup Analysis of Nonfatal Myocardial Infarction Endpoint

For subgroup analyses all adjudicated nonfatal MI outcomes will be included, and the analyses will be performed for the ITT population. The potentially clinically relevant baseline risk factors and medication use for exploratory subgroup analyses for the MI events will include: prior MI, prior hospitalization for unstable angina, prior coronary revascularization, current smoker, hypertension, hyperlipidemia, aspirin use, beta-blocker use, calcium channel blocker use, ACE inhibitor or angiotensin receptor blocker use, diuretic use, and antithrombotic use (excluding aspirin).

The risk-adjusted analysis will be performed using these risk factors, medication use, the prespecified subgroups (Section 4.13.3), and a multivariate proportional hazards model. The risk factors and medication use remaining in the final model plus the pre-specified subgroup variables will be used for subgroup analyses.

4.14.2.3. Analysis of Fatal Myocardial Infarction Endpoint

A fatal MI is defined as an MI event that resulted in death within 30 days. Time-to-event analysis (Section 4.3.4) will be performed for all adjudicated fatal MI outcomes using the ITT population. Subgroup analyses will also be performed if there are enough cases of fatal MIs (Section 4.3.5 and Section 4.13.3).

4.14.3. Additional Analyses of All Stroke Endpoint

Time-to-event analyses will be performed for each subtype of stroke (hemorrhagic, ischemic, or unknown) in the ITT and PP populations. Incidence rates per 100 person-years of follow-up will be calculated for each type of stroke for each treatment group. The NNT statistic will not be provided for these analyses.

Stroke disability, as measured using the modified Rankin scale (Farrell et al. 1991), will be assessed at approximately 30 days after the diagnosis. The number of patients in each of the following categories of the scale will be summarized and compared between treatment groups using a Chi-square test Fisher's exact test:

0: No symptoms at all

1: No significant disability despite symptoms (Able to carry out all usual activities)

2: Slight disability

3: Moderate disability (Requiring some help but able to walk without assistance)

4: Moderate to severe disability (Unable to walk without assistance and unable to attend to own bodily needs without assistance)

5: Severe disability (Bedridden, incontinent and requiring constant nursing care and attention)

6: Death.

4.14.4. Subgroup Analysis of Nonfatal Stroke Endpoint

For subgroup analyses all adjudicated nonfatal stroke outcomes will be included, and the analyses will be performed for the ITT population. The potentially clinically relevant baseline risk factors and medication use for exploratory subgroup analyses for the stroke events will include: prior MI, prior hospitalization for unstable angina, prior history of stroke or TIA, carotid or other artery disease, current smoker, alcohol abuse, hypertension, hyperlipidemia, aspirin use, beta-blocker use, calcium channel blocker use, ACE inhibitor or angiotensin receptor blocker use, diuretic use, history of atrial fibrillation, and antithrombotic use (excluding aspirin).

The risk-adjusted analysis will be performed using these risk factors and the prespecified subgroups, and a multivariate proportional hazards model. The risk factors and medication use remaining in the final model plus the prespecified subgroup variables will be used for subgroup analyses.

4.14.5. Analysis of Fatal Stroke Endpoint

A fatal stroke is defined as a stroke event that resulted in death within 30 days. Time-to-event analysis (Section 4.3.4) will be performed for all adjudicated fatal stroke outcomes using the ITT population. Subgroup analyses will also be performed if there are enough cases of fatal strokes (Section 4.3.5 and Section 4.13.3).

4.14.6. Exploratory Analyses: Analyses of Multiple Cardiovascular Events

A patient may experience more than one CV event during the course of the study. For example, a patient may be hospitalized for an unstable angina and later experience a nonfatal MI which

may be followed by death from a CV cause, or a patient may experience 3 nonfatal MIs or 2 nonfatal strokes. In each case, the patients will be considered as having multiple CV events. To assess the effects of dulaglutide on multiple CV events compared to placebo, analyses accounting for multiple CV events will be performed. The analyses will be performed separately for the composite primary endpoint; acute coronary syndrome (ACS) events which include nonfatal MI and hospitalization for unstable angina; MIs; strokes; MACE (composite of nonfatal MI, stroke and CV death) and heart failures requiring hospitalization. In each of these analyses, a patient who experiences multiple events of the endpoint will be considered as having recurrent events of the endpoint. The conditional Gap Time model of Prentice, Williams, and Peterson (PWP-GT) (Prentice et al. 1981) will be used to analyze recurrent CV events. The hazard ratio with 95% CI will be reported for time to the first event, second event, etc. The mean time between successive events will also be reported by treatment group. Baseline characteristics will be compared between groups of patients with no event, only one event, and multiple events (2 or more) regardless of treatment assignment and between treatments within each group. Counts and proportion for categorical variables will be compared using a Chi-square or Fisher's exact test and means for continuous variables will be compared using ANOVA.

4.14.7. Exploratory Analyses: Analysis of Nonfatal Cardiovascular Events with Competing Risk of Death

The composite endpoint of nonfatal CV events (that is, the first occurrence of nonfatal MI, nonfatal stroke, hospitalization for unstable angina, HF requiring hospitalization or an urgent HF visit, or coronary, carotid, or peripheral revascularization), will be analyzed taking into account the competing risk of death. The competing risk model of Fine and Gray will be used (Fine and Gray 1999). Covariates of interest include treatment and the baseline CV risk factors listed in Section 4.13.3.

4.15. Additional Measures

Additional measures include:

- Change from baseline in:
    - hemoglobin A1c levels
    - weight
    - waist/hip ratio
    - cognitive function as measured by the MoCA and the DSST
    - erectile dysfunction as measured by the IIEF
- Time to first occurrence of (after randomization):
    - composite endpoint of death from CV causes, nonfatal MI, nonfatal stroke, or hospitalization for unstable angina

- o coronary, carotid, or peripheral revascularization, individually and compositely

- o any hospitalization

- Incidence of:

    - o any fracture

    - o development of cholelithiasis

An independent CEC will adjudicate coronary, carotid, and peripheral revascularizations. Detailed event definitions and specifics regarding endpoint determination are provided in the CEC Charter.

- A **coronary, carotid, or peripheral arterial revascularization** procedure will be defined as a catheter based or open surgical procedure designed to improve myocardial, carotid, or peripheral arterial blood flow. Insertion of a guidewire through a coronary guide catheter into a coronary artery or bypass graft for the purpose of PCI is considered intention for PCI. The intention to perform percutaneous peripheral arterial intervention is denoted by the insertion of a guidewire through a guide catheter into a peripheral artery.

- A **hospitalization** will be defined as a hospital admission (including admission to a chest pain observation unit) or a visit to an emergency department that results in a stay >24 hours.

- A **fracture** will be defined as a clinically or radiologically apparent fracture of any bone.

- **Development of Cholelithiasis** will be defined as any new diagnosis of cholelithiasis after randomization, as evidenced on an imaging examination (for example, ultrasound or computerized tomography scan).

4.15.1. Analyses of HbA1c, Weight, and Waist/Hip Ratio

ForHbA1c, weight, and waist/hip ratio, an MMRM analysis for the change from baseline to each visit will be performed for the ITT population (Section 4.3.2). Baseline for HbA1c and weight will be Visit 1 (Screening). A correction factor will be used to standardize the HbA1c values reported depending on whether the measurement method was IFCC or DCCT.

4.15.2. Analyses of Cognitive Function and Erectile Dysfunction

Cognitive function will be assessed in all randomized patients using the MoCA instrument and the DSST at baseline (Visit 3) and after 24, and 60 months of treatment and at the final visit.

The MoCA is a cognitive screening test designed to detect mild cognitive impairment (Nasreddine et al. 2005). It assesses different cognitive domains: attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking,

calculations, and orientation. It will take approximately 10 minutes to complete the test. The total possible score is 30 points; a score of 26 or above is considered normal.

The DSST is an attention-demanding psychomotor component of the Wechsler Adult Intelligence Scale (Kuo et al. 2007). The patient is given a symbol-digit code in which each of the digits 1 through 9 is paired with a different symbol. Below the code, a series of symbols selected from those in the code are presented in an irregular order. The patient is instructed to draw the symbol that matches the number and to complete as many correct symbols as possible within a 120-second test period. The DSST score is the number of correct symbol-number matches. The number of matches attempted will also be recorded.

Analyses of the last score for the MoCA and the DSST, and visit-specific analyses will be performed using MMRM (Section 4.3.2) for each of these continuous test scores. The analysis will be based on change from baseline. Patients will be required to have a baseline and at least one post-baseline score to be included in these analyses.

The MoCA score is a continuous variable with a range of [0, 30]. It will also be analyzed as a categorical variable using the categories: below the threshold for normal cognitive function (that is, mild cognitive dysfunction, MoCA score <23 for Korea and MoCA score <26 for all other countries), and above the evaluation threshold (that is, normal cognitive function, MoCA score ≥23 for Korea and MoCA score ≥26 for all other countries). The number and proportion of patients in each category at each visit will be compared between treatment group using a Chi-square test or Fisher's exact test.

4.15.3. Analyses of Erectile Dysfunction

Another additional objective is to evaluate the effect of dulaglutide compared to placebo on the IIEF scores in men. This objective will be assessed using the measure of change from baseline to each visit in total IIEF scores from the 15-item questionnaire in the erectile function (EF), orgasmic function, sexual desire, overall satisfaction (OS), and intercourse satisfaction (IS) domains. The analysis will use an MMRM model that includes among others (Section 4.3.2) terms for treatment and the baselines values minus their mean as covariate. The population for these analyses will be all randomized male patients.

4.15.4. Analyses of Other Additional Measures

Time-to-event analyses will be performed for each of the following endpoints: the composite endpoint of death from CV causes, non-fatal MI, nonfatal stroke, or hospitalization for unstable angina; the composite endpoint of coronary, carotid, or peripheral revascularization and each of the components; any hospitalization; any fracture; and the development of cholelithiasis. A Cox proportional hazards regression model for the time to the first occurrence of the event, with treatment as fixed effects will be performed for the ITT population. Kaplan-Meier curves, hazard ratios and associated 95% CIs will be provided. Treatment groups will be compared using the p-value from the Cox model. The population for the analysis of cholelithiasis will be all randomized patients who have not had a cholecystectomy before randomization.

Revascularizations are classified as elective or nonelective and as successful or not successful. The number and proportion of revascularizations falling in each category will be compared between treatment groups using a Chi-square or Fisher's exact tests. Reasons for hospitalization (adverse Event, endpoint, or other) will also be compared between treatment groups.

4.16. Other Exploratory Analyses

4.16.1. Analyses of Time to Glycemic Intervention

Additional therapeutic intervention may be considered (with the exception of a GLP-1 analog or pramlintide) in patients who do not attain target HbA1c values and/or develop severe hyperglycemia, despite full compliance with the assigned study treatment regimen. These changes may be instituted 3 months after randomization to enable the effects of study drug on HbA1c to stabilize, unless sooner intervention is indicated, in the judgment of the investigator.

According to the protocol, patients should continue to inject their allocated study drug and will remain in the study.

Time-to-event analyses will be performed for the time to the first glycemic intervention after randomization. The incidence of glycemic interventions will be summarized using number and proportion of patients by treatment group and by visit. The overall number and proportion will also be reported as will, Kaplan-Meier estimates of the proportion of patients having 1 or more glycemic intervention by treatment group. The number and proportion at each visit are calculated as the number of patients and proportion of patients reporting glycemic interventions at that visit. The overall number and proportion are calculated as the total number of patients and proportion of patients reporting severe glycemic interventions during the entire study treatment period. Treatment group comparison will be assessed using a Chi-square or Fisher's exact tests or log-rank tests as appropriate. The mean time to initiation of additional therapies will be compared between treatment group using an ANOVA model with treatment and additional covariates such baseline HbA1c and concomitant antihyperglycemic agents.

4.16.2. Analyses Stratified by Baseline Concomitant Medications

Exploratory time-to-event analyses of the primary CV endpoint will be performed stratified by the following categories of concomitant medications taken by randomized patients at baseline: antihyperlipidemic agents (Yes/No), antihypertensive agents (Yes/No), and antithrombotic agents (Yes/No). The analyses will be performed using the ITT population.

4.17. Safety Analyses

Unless otherwise noted, all listings and all summary analyses will be conducted using all randomized patients (that is, ITT population). The routine safety analyses will include the measurements of treatment emergent adverse events (TEAEs), serious adverse events (SAEs), laboratory analytes, vital signs, and ECGs.

4.17.1. Prespecified Safety Measures (Adverse Events of Special Interest)

Prespecified safety measures include the incidence of:

- acute pancreatitis

- serious GI events

- any cancer (excluding basal or squamous cell skin cancer) and specific categories of

  o pancreatic cancer

  o medullary thyroid carcinoma (MTC) and C-cell hyperplasia

  o thyroid carcinomas

- severe hypoglycemia

- immune mediated reactions including serious allergic and hypersensitivity reactions

- serious hepatic events

- clinically significant supraventricular arrhythmias and cardiovascular conduction disorders

- serious renal events

- discontinuation of study drug for any reason

4.17.1.1. Analysis of Severe Hypoglycemia

Severe hypoglycemic episodes by patient by visit will be listed using all randomized patients.

The incidence of severe hypoglycemic episodes will be summarized using number and proportion by treatment group and by visit. The overall number and proportion will also be reported as will Kaplan-Meier estimates of the proportion of patients having 1 or more events by treatment group. The number and proportion at each visit are calculated as the number of patients and proportion of patients reporting severe hypoglycemic episodes at that visit. The overall number and proportion are calculated as the total number of patients and proportion of patients reporting severe hypoglycemic episodes during the entire study treatment period. Treatment group comparison will be assessed using a Chi-square or Fisher's exact tests or log-rank tests as appropriate.

Severe hypoglycemia rate per year (number of events per subject per year) will be summarized by yearly visit by treatment group. The rate will be analyzed if enough data points are available. The rate of hypoglycemia will be analyzed using a generalized mixed effect model with a negative binomial distribution (see SAS code below).

proc glimmix data=data_glm noitprint empirical initglm method=rspl;

nloptions maxiter=**100** tech=NRRIDG;

class subjid trt visit country stratification variables ;

model count= baseline count country stratification variables trt visit trt*visit/dist=nb
 link=log offset=log_days s ddfm=kr;

random visit/subject= subjid type=un residual;

run;

An unstructured covariance structure will be used to model the within-patient errors. If this analysis fails to converge, the following covariance structures will be tested in this order: compound symmetry, then, autoregressive. The model will include pooled country, treatment, visit, visit-by-treatment interaction, and baseline. Other covariates of interest, including categorical, continuous and time-dependent may be included.

Time to the first severe hypoglycemia event per patient will also be analyzed via a Cox model described in Section 4.3.4. Incidence rates for 100 person-years will also be presented.

4.17.1.2. Analysis of Other Prespecified Safety Measures

Each of the following events will be analyzed using the ITT population: acute pancreatitis, serious GI events, any cancer (excluding basal or squamous cell skin cancer) and specific categories of pancreatic cancer, MTC and C-cell hyperplasia, and thyroid carcinomas, serious hepatic events, clinically significant supraventricular arrhythmias and cardiovascular conduction disorders, serious renal events, discontinuation of study drug for any reason, and immune mediated reactions including serious allergic and hypersensitivity reactions (overall and by type: rash, urticaria, bronchospasm, angioedema, systemic anaphylaxis, erythema multiforme, Stevens Johnson syndrome, toxic epidermal necrolysis, and other). The timing of the allergic/hypersensitivity reactions (immediate reaction - occurs within minutes [<60 minutes], acute reaction - occurs from 1 up to 6 hours from study drug administration, >6 hours through 7 days from study drug administration, and >7 days from study drug administration) will also be summarized and compared between treatment groups. Acute pancreatitis will be analyzed based on adjudicated events and on events as reported by investigators. The analyses of medullary thyroid carcinoma, other thyroid carcinomas, and C-cell hyperplasia will be based on adjudicated events. The analyses of clinically significant supraventricular arrhythmias and cardiovascular conduction disorders and immune mediated reactions including serious allergic and hypersensitivity reactions will be based on SAE reports. The analysis of other cancers (excluding basal or squamous cell skin cancer) will be based on events reported by investigators.

The incidence will be summarized using number and proportion by treatment group for each of the prespecified safety events. The overall number and proportion will be reported. The overall number and proportion will be calculated as the total number of patients and proportion of patients reporting the event during the entire study treatment period. In addition, incidence rates per 100 person-years will be calculated for each of the prespecifed safety events. Treatment group comparison will be assessed using a Chi-square or Fisher's exact tests.

4.17.2. Adverse Events

A patient's safety reporting period will commence on the date of randomization to the time of last participation in the study (Section 4.12).

An AE is defined as any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and that does not necessarily have a causal relationship with this treatment. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the medicinal (investigational) product.

Adverse events will be classified according to the Medical Dictionary for Regulatory Activities (MedDRA™). Analysis of AEs will focus primarily on those events that first occur or worsen (increase in frequency or severity) after the first injection of study drug following randomization, (that is, TEAEs); however, all conditions/events reported on the PRE-EXISTING CONDITIONS AND ADVERSE EVENTS page of the case report form (CRF) will be retained in the study database and will be reported in the listings. Study drug overdose will also be reported as a TEAE. Study drug overdose is defined as documented evidence of study drug injection more than once in a 3-day period. Analyses of TEAEs will be based on all randomized patients.

The identification of TEAEs will be performed according to the following process. Each condition/event will be coded to a preferred term (PT). For each patient, events will be divided into 2 groups: baseline events will include all events present prior to the date of randomization and before the first injection of study drug post randomization, and post-baseline events will include all events present on or after the date of randomization and the date of first injection of study drug. The maximum severity level (mild, moderate, or severe) reported prior to randomization and the first injection of study drug among all baseline adverse events that code to the same lower level term will be reported. If severity is missing for a particular baseline adverse event, then "mild" will be assumed. For each post-baseline event, the maximum severity reported after randomization and the first injection of study drug will be compared with the maximum baseline severity for the corresponding lower level term. If severity is missing for a particular postbaseline adverse event, then "severe" will be assumed. When the maximum post-baseline severity exceeds the maximum baseline severity, the event will be classified as a TEAE.

The primary hierarchy (PT through System Organ Class [SOC]) associated with the lower level term will be assigned. The number of patients experiencing a TEAE will be compared between treatments at the PT and SOC levels.

A by-patient listing of all adverse events, treatment-emergent or not, will be provided.

4.17.2.1. Serious Adverse Events

A SAE is any AE from this study that results in one of the following outcomes (please note exceptions outlined below):

- death (except as noted below)

- initial or prolonged inpatient hospitalization

- a life-threatening experience (that is, immediate risk of dying)

- persistent or significant disability/incapacity

- congenital anomaly/birth defect

- considered significant by the investigator for any other reason.

The following primary, secondary, and additional efficacy events will not be required to be reported as AEs or SAEs *unless* the investigator believes the event may have been caused by the study drug, drug delivery system, or study procedure:

- death

- nonfatal MI

- nonfatal stroke

- hospitalization for HF or an urgent HF visit

- hospitalization for unstable angina

- coronary, carotid, or peripheral revascularizations

If one of the above endpoint events is reported but does not meet a prespecified event definition detailed in the CEC Charter, as reviewed by the independent CEC, the study site subsequently will be required to report the event as an AE or SAE to comply with regulatory reporting requirements.

All SAEs will be compared between treatment groups at the PT level in the ITT population.

4.17.2.2. Treatment-Emergent Adverse Events Leading to Permanent Discontinuation of Study Drug or Study Withdrawal

Patients may permanently discontinue study medication but are expected to remain in the study. Separate analyses will be performed to compare the treatment groups with respect to AEs that led to permanent discontinuation of study medication and AEs that led to withdrawal from the study. These analyses will be conducted for all randomized patients.

4.17.3. Laboratory Analytes

All laboratory measurements will be performed locally except calcitonin which will be performed by a central laboratory. For analyses, the laboratory measurements will be converted to SI units. The following laboratory measurements will be analyzed: calcitonin, creatinine, eGFR, serum creatinine, urine ACR, HbA1c (Section 4.15.1) and lipids (Section 4.17.3.1).

Laboratory measurements collected at scheduled and unscheduled visits will be listed by patient by visit using all randomized patients. An additional listing will be presented for all scheduled and unscheduled laboratory measurements that are outside the SI normal range. All summary

analyses will be based on the ITT population. Laboratory measurements that fall within a visit window (±30 days around the observed visit date) will be associated with that visit. Scheduled visits will have a ±15-day window around the scheduled visit date. For multiple runs of the same laboratory analyte, the laboratory measurement within the window that was taken closest to the visit date will be representative of that patient's lab value for that visit.

The eGFR values will be calculated using the MDRD equation [eGFR (mL/min/1.73 m^2) = 175 X standardized Scr $^{-1.154}$Xage $^{-0.203}$ X1.212 [if black] X 0.742 [if female], (Levey et al. 2006)].

Section 4.2 describes the analysis for continuous laboratory measurements. Microalbuminuria is defined as a urine ACR >30 mg/g and ≤300 mg/g) and macroalbuminuria (clinical proteinuria) as urine ACR >300 mg/g). Urine ACR is calculated as the ratio of urine albumin and creatinine measured from a morning urine sample or a random urine sample if a morning sample is not available.

The proportion of patients with microalbuminuria or macroalbuminuria at baseline or during the study will be compared between treatment groups at by visit. Shift tables will be presented for urine ACR. The tables will show the proportions of patients with shifts in the laboratory results from baseline to maximum post-baseline result using categories based on the Central Laboratory reference ranges. The categories for urine ACR will be ≤ 30 mg/g, >30 mg/g to ≤300 mg/g and >300mg/g. Shifts will be grouped to show the proportions of patients who experienced decreases (post-baseline category < baseline category), increases (post-baseline category > baseline category), or no change from baseline to maximum post-baseline result for each treatment. Baseline will be the last non-missing observation in the study period from screening to randomization (Visit 1 through Visit 3). The maximum post-baseline result in the postbaseline study period (Visits 3 and beyond) will be the maximum result for the analysis. The analyses will use both scheduled and unscheduled labs and will be in all randomized patients with at least 1 baseline and 1 post-baseline urine ACR measurement.

4.17.3.1. Lipid Parameters
The following lipid parameters will be assessed for all randomized patients:

- total cholesterol

- high-density lipoprotein cholesterol (HDL-C)

- low-density lipoprotein cholesterol (LDL-C)

- non-HDL-C (total cholesterol minus HDL-C)

- ratio of total cholesterol to HDL-C (total cholesterol divided by HDL-C)

- triglycerides

Fasting lipids will be collected at randomization (Visit 3) and after 24 and 60 months of treatment (Visit 15). Visit 3 will be considered as baseline for these analyses. Analyses of the percent change from baseline to each visit and to last measurement will be performed separately

for each marker. Analyses of change from baseline to each visit and to last measurement will be performed for the ratio of total cholesterol to HDL-C. The median will be reported for triglycerides since the data are not expected to be normally distributed; and the mean will be reported for the other lipid parameters. Patients will be required to have a baseline and a postbaseline measurement of the marker to be included in these analyses.

4.17.4. Vital Signs

Vital signs (systolic blood pressure, diastolic blood pressure, and heart rate) will be collected in the seated position in triplicate at each office visit. Measurements will be averaged for each patient at each visit; the average values will be used in the descriptive summaries and analyses.

All averaged measurements will be listed by patient by visit using all randomized patients.

Descriptive statistics for the averaged measurements and change from baseline by treatment arm and visit will be presented. Summary analyses will be conducted using the ITT population. The change from baseline will be analyzed using the MMRM model.

4.17.5. ECG Analyses

Electrocardiograms will be performed for all randomized patients at each visit. Both scheduled and unscheduled ECGs will be qualitatively evaluated by an ECG reading center. The qualitative characteristics assessed will be summarized in the major categories and subcategories of findings: normal ECG, abnormal ECG findings, and the subcategories of abnormal findings.

The number of patients in each category and subcategory will be compared between treatment groups and by visit using a Chi-square or Fisher's exact tests.

5. Unblinding Plan

5.1. Introduction

This unblinding plan describes the organization of personnel and definition of processes that will be followed to insure integrity of the data and results throughout the trial.

5.2. Organization of REWIND Study

The Steering Committee (SC) will be responsible for the overall scientific conduct of the study and all scientific trial-related decisions, and will assist with local issues to support the implementation and good conduct of the study worldwide.

Lilly will assign the obligation of study operation management to a CRO. ICON will be the CRO for this study. Medical oversight will be the responsibility of Lilly and the CRO. The CRO will be responsible for addressing medical and study operational questions, handling the data, conducting the analyses, producing all data summaries for the CSR as well as writing the CSR.

An IDMC will be responsible for monitoring patient safety and will review unblinded interim and safety analyses during the study. An ISAC will perform analyses for the IDMC prior to unblinding. The members of the ISAC are employees of Population Health Research Institute (PHRI).

For this unblinding plan, REWIND personnel will fall into one of the following 2 categories:

Personnel blinded until datalock:

- investigators

- patients

- Steering Committee members*

- ICON study personnel

- Lilly personnel* except a limited number of Lilly IVRS and clinical trial material representatives

- Population Health Research Institute (PHRI) study personnel except those who are part of the ISAC

  *There may be a rare exception to the blinded status of the SC Chair and Lilly personnel as noted in the IDMC charter.

At ICON, a blinded statistical team will be responsible for the production and quality assurance review of analysis datasets and their periodic transfers to Lilly and the ISAC.

Until the final database lock, no unblinded reports will be accessible to study personnel who are to remain blinded until the end of the trial.

Personnel unblinded from the time of the first unblinded analysis:

- IDMC members

- ISAC

- Lilly IVRS and clinical trial material representatives

Depending on the recommendation of the IDMC, *a* Lilly internal review group (IRG) may be unblinded to study data. The Lilly IRG comprises a limited number of internal Lilly medical and statistical experts, who have no direct involvement in the clinical development of dulaglutide.

To preserve the blinding of the study, a minimum number of Lilly personnel will see the randomization table and treatment assignments before the study is complete. This controlled access will be limited to certain IVRS or clinical trial material personnel. The IVRS representative will not be responsible for any analysis decisions or production work for the IDMC summaries and analyses that are described in the IDMC charter. The ISAC is responsible for creating and reviewing test programs and outputs for IDMC reports based on data blinded at the treatment group level (Treatment A versus Treatment B) prior to each IDMC review. The ISAC statistician is responsible for the production of the unblinded summaries and analyses for the IDMC from programs developed by the ISAC described above.

5.3. Data Handling and Storage

Transfers of data to the ISAC and to Lilly will be the responsibility of the blinded statistician and data management group at the CRO. Further details on data handling will be provided by the CRO.

The Lilly IVRS team will generate and maintain the randomization schedule and treatment group assignments. All treatment code information will be stored electronically on a secure server system directory with access only to the Lilly IVRS team members. This team will provide the treatment codes to the unblinded ISAC statistician through secure router ID at the time of the analyses for the IDMC.

5.4. Unblinded Data

Access to unblinded data by treatment group will be restricted to the IDMC and ISAC until the trial reaches its scheduled termination. If the IDMC recommends early termination, unblinded data will be provided to the sponsor. If the IDMC recommends a major protocol change, the IDMC Chair, the Lilly Research Laboratories Designee, and the Chair of the Steering Committee will discuss whether the Lilly Internal Review group or the Chair of the Steering Committee, or both, should review select unblinded data; the data to be reviewed will depend on the issues raised.

The IDMC understands that Lilly will behave in accordance with its own internal policies regarding review of unblinded data.

The IDMC may request, in a confidential manner, input from additional independent scientists to assist in the IDMC's decision-making. The IDMC may share select blinded data from the trial with these scientists.

5.5. Other Unblinding Issues

5.5.1. Unblinding of Individual Patients during the Study

To preserve the blinding of the study, a minimum number of Lilly personnel will see the randomization table and treatment assignments before the study is complete. This controlled access will be limited to certain IVRS or clinical trial material personnel. However, all personnel involved with the study, including the SC, all investigators, all Lilly personnel (excluding those referenced above) and all CRO personnel, and anyone other than those people charged with assuring the safety of the trial (such as, the IDMC) and drug will be blinded to all post-randomization data by treatment group.

Emergency unblinding for AEs may be performed through IVRS. This option may be used ONLY if the patient's well-being requires knowledge of the patient's treatment assignment. All calls resulting in an unblinding event are recorded and reported by the IVRS.

Study site personnel must alert the CRO within 1 business day of the investigator unblinding a patient's treatment group assignment for any reason.

Lilly Global Patient Safety and CRO will review SAEs within time frames mandated by company procedures. If a death or clinical AE is deemed serious, unexpected, and possibly related to study drug, Lilly Global Patient Safety and CRO will be unblinded to comply with regulatory reporting and safety monitoring requirements. These measures will preserve the integrity of the data collected during this study and minimize any potential for bias while providing for appropriate safety monitoring.

In the event that safety monitoring uncovers an issue that needs to be addressed by unblinding at the group level, only members of the IDMC and the ISAC, that provides support to the IDMC, can view group unblinded data and conduct additional analyses of the safety data.

The unblinded ISAC statistician will monitor the number of patients unblinded during the study for any trends that need to be raised to the attention of the IDMC Chair. The ISAC statistician will also maintain a list of people who received unblinded information during the trial.

Periodically and at the end of the study, the Lilly IVRS representative will provide the details of the patients and the date that they were unblinded to the unblinded ISAC statistician for review by the IDMC. Following the final review before the trial ends, the unblinded ISAC statistician will place this information into the correct server directory with all the other interim reports to be sent to the Lilly blinded statistician at the end of the trial.

5.5.2. Other Data That May Unblind Patients

It is assumed that there is no other clinical data (for example, laboratory results, ECG findings, vital sign measurements) that might potentially disclose individual patient treatment assignment to investigators, the Steering Committee, or all other Lilly and ICON study personnel during the conduct of the study.

6.　References

Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med*. 2014;33(4):693-713.

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med*. 2009;28(4):586-604.

Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J*. 2011;53(6):894-913.

Farrell B, Godwin J, Richards S, Warlow C. The United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: final results. *J Neurol Neurosurg Psychiatry*. 1991;54(12):1044–1054.

Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496–509.

Jennison C, Turnbull BW. Group sequential methods with applications to clinical trials. Boca Raton, FL:Chapman and Hall/CRC; 2000.

Kuo HK, Leveille SG, Yu YH, Milberg WP. Cognitive function, habitual gait speed, and late-life disability in the National Health and Nutrition Examination Survey (NHANES) 1999-2002. *Gerontology*. 2007;53(2):102-110.

Levey AS, Coresh J, Greene T, Stevens LA, Zhang YL, Hendriksen S, Kusek JW, Van Lente F; 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(4):247-254.

Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699.

O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979;35(3):549-556.

Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika*. 1981;68(2):373–379.

Thygesen K, Alpert JS, White HD. Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. *Eur Heart J*. 2007;28(20):2525-2538.

Whitehead J. Overrunning and underrunning in sequential clinical trials. *Control Clin Trials*. 1992;13(2):106-121.